

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-1

Perfect score: 15

Sequence: 1 acctcttgatcgcgc 15

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_btg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	DB		
1	15	100.0	15	6	ARI83474	ARI83474 Sequence
2	15	100.0	286	9	HSYHLS17	273860 H.sapiens r
3	15	100.0	15068	4	MYRIR2	y07749 M.vision MRN
4	14	93.3	1018	11	CNS06JG2	ALA01544 t7 end of
5	14	93.3	1443	6	E15997	E15997 cDNA encodi
6	14	93.3	1626	6	E15998	E15998 cDNA encodi
7	14	93.3	1638	6	HSFBRUP2	Y16280 Homo sapien
8	14	93.3	2411	6	AX018426	AX018426 Sequence
9	14	93.3	10585	1	AE001342	AE001342 Chlamydia
10	14	93.3	16360	4	MAC03116	AJ303116 Tachyglis
11	14	93.3	55739	8	AP000389	AP000389 Arabidops
12	14	93.3	67079	2	AC123621	AC123621 Mus muscu
13	14	93.3	68054	9	AL592300	AL592300 Human DNA
14	14	93.3	69928	2	AC100880	AC100880 Mus muscu
15	14	93.3	73509	2	AL137862	AL137862 Homo sapi
16	14	93.3	83605	2	AC111830	AC111830 Rattus no
17	14	93.3	100406	2	AC122585	AC122585 Rattus no
18	14	93.3	123016	2	AF189001	AF189001 Homo sapi
19	14	93.3	125703	2	AC123338	AC123338 Rattus no
20	14	93.3	160764	9	AC107928	AC107928 Homo sapi
21	14	93.3	161618	2	AC122459	AC122459 Mus muscu
22	14	93.3	168346	2	AC036227	AC036227 Homo sapi
23	14	93.3	172862	2	AL645796	AL645796 Homo sapi
24	14	93.3	176593	2	AC087661	AC087661 Homo sapi
25	14	93.3	179901	2	AC011006	AC011006 Homo sapi
26	14	93.3	181672	9	AC079822	AC079822 Homo sapi
27	14	93.3	187134	2	AC121851	AC121851 Mus muscu
28	14	93.3	189608	9	AL359253	AL359253 Human DNA
29	14	93.3	199650	9	AF252829	AF252829 Homo sapi
30	14	93.3	199810	2	AC091562	AC091562 Homo sapi
31	14	93.3	200355	2	AC019218	AC019218 Homo sapi
32	14	93.3	226715	2	AL596104	AL596104 Mus muscu
33	14	93.3	259854	2	AC098462	AC098462 Rattus no
34	14	93.3	281659	2	AC074159	AC074159 Mus muscu
35	13.4	89.3	39	6	ARI16990	ARI16990 Sequence
36	13.4	89.3	39	6	ARI16991	ARI16991 Sequence
37	13.4	89.3	60	6	168042	168042 Sequence 11
38	13.4	89.3	217	9	HS52A9R	Z61427 H.sapiens C
39	13.4	89.3	263	9	HS93B1R	Z63943 H.sapiens C
40	13.4	89.3	265	9	HS8F1B01	Z75463 H.sapiens r
41	13.4	89.3	341	5	AF271485	AF271485 Salmo sal
42	13.4	89.3	360	6	181217	181217 Sequence 10
43	13.4	89.3	363	9	U00528	U00528 Human immu
44	13.4	89.3	366	6	AX112690	AX112690 Sequence
45	13.4	89.3	372	9	HUMIGHCY	M34678 Human Ig ac

ALIGNMENTS

RESULT 1

LOCUS ARI83474 15 bp DNA linear PAT 20-APR-2002

DEFINITION Sequence 1 from patent US 6342220.

ACCESSION ARI83474

VERSION ARI83474.1 GI:20227443

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unknown.

REFERENCE 1 (bases 1 to 15)

AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.

TITLE Agonist antibodies

JOURNAL Patent: US 6342220-A 1 29-JAN-2002;

FEATURES Location/Qualifiers

```

source          1..15
                /organism="unknown"
BASE COUNT      2 a      5 c      4 g      4 t
ORIGIN
Query Match     100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTGGATCGGC 15
    |||||||
    1 ACCTCTGGATCGGC 15

RESULT 2
LOCUS HSVHL517
DEFINITION H.sapiens rearranged DNA for Ig heavy chain, VH251-D-J, clone
L5-17.
ACCESSION Z73860.1 GI:1495558
VERSION Z73860.1 GI:1495558
KEYWORDS diversity region; immunoglobulin heavy chain; immunoglobulin
superfamily; joining region; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 286)
AUTHORS Dunn-Walters,D.K.
JOURNAL Direct Submission
Submitted (14-MAY-1996) Dunn-Walters D.K., UMDS St Thomas' Campus,
Dept of Histopathology, Lambeth Palace Road, London SE1 7EH, UK
2 (bases 1 to 286)
AUTHORS Dunn-Walters,D.K. and Spencer,J.
JOURNAL Origins of lamina propria plasma cells
Proc. Ninth Int. Congress for Mucosal Immunol.
FEATURES
    source
        1..286
            /organism="Homo sapiens"
            /isolate="specimen B-2, area 31"
            /db_xref="taxon:9606"
            /chromosome="14"
            /clone="L5-17"
            /cell_type="plasma cell"
            /tissue_type="ileal lamina propria"
            /rearranged
        D-segment 241..273
        J-segment 274..286
        BASE COUNT 64 a      89 c      69 g      64 t
        ORIGIN
Query Match     100.0%; Score 15; DB 9; Length 286;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTGGATCGGC 15
    |||||||
    37 ACCTCTGGATCGGC 51

RESULT 3
LOCUS MYR3
DEFINITION M.vision mRNA for ryanodine receptor 3.
ACCESSION Y07749.1 GI:1526614
VERSION Y07749.1 GI:1526614
KEYWORDS ryanodine receptor 3; RYR3 gene.
SOURCE Mustela vison.
ORGANISM Mustela vison.
REFERENCE 1 (bases 1 to 15068)

```

```

AUTHORS Martzali,G., Rossi,D., Giannini,G., Charlesworth,A. and
          Sorrentino,V.
TITLE cDNA cloning reveals a tissue specific expression of alternatively
          spliced transcripts of the ryanodine receptor type 3 (RYR3) calcium
          release channel
JOURNAL 2 (bases 1 to 15068)
REFERENCE 2 (bases 1 to 15068)
AUTHORS Sorrentino,V.
JOURNAL Direct Submission
Submitted (02-SEP-1996) V. Sorrentino, DIBIT, Istituto Scientifico
          San Raffaele, via Olgettina 58, I-20132, Milano, ITALY
FEATURES
    source
        1..15068
            /organism="Mustela vison"
            /db_xref="taxon:9667"
            /cell_line="CCL64"
            1..14580
            /gene="RYR3"
            1..14580
            /gene="RYR3"
            /codon_start=-1
            /product="ryanodine receptor type 3"
            /protein_id="CA669029.1"
            /db_xref="GI:1526615"
            /db_xref="SPTREMBL:095201"
            /translation="MARGGEGDELEIOLPTEDEYVLOCIANIHKREKRECLAEGLG
            NLFCELEPTESEAKVTPDLCVNFVLEOSLSVVALQDEMLANGENGEGEAGGGHRT
            LLVGHALLRHSPSGMYLTCTLTSTRSQTOKLAVDGLRNATNGACGMWTHIRASQRS
            EGKVRVIRGDDLLIVSVSEERYLLHSVNGWVQVADSFQMTLNVPTSGSSITEGYL
            LGCHVVRFLFHGHDECLIPSTDONDAOHRIEFAAGASARARLVRVDSISGSL
            NRMGOAFRLRHITTYGHLATDEOGLLLODRKSPDKSTKSPKRSKIKRKLHDEBMD
            KRDEMGVPEIRKIGDVCVFOHJAGLWVYAKOAKRSTKSPKRSKIKRKLHDEBMD
            DCLTORCOREESQARIIRNTALFSPQSGNRRAAVPTPIQEVLOTODLIATF
            OPPEEMQHEDEKOKRLSLKRONLREKEMALVYNCIDRINTNSAHRFGLTARE
            SGMAKEDILNLTALIRGNKNCAGSNLDMVLSKTRLESSGTELEVCITL
            IESPEALNLAEGHISISLIDKGNHKNVLDVLCISLNCGVAVRANQNLICNLL
            PRNLLIOTRLINDVTSIRENIFLGAESGAVQKKVFEILLIDVDFLEATPTLAV
            GMASSGYAPYPGGEGMGNGYGDLDYSGEPDGLMAGRIPIRAVASINHLIRSD
            VVSCICDGVPSISFRINGOPVQGMFENETDLPFPVVSFAGVAVRLTGTGRGEF
            KLPSPGAVACYEALLPKEKMLPEVKEKRRDGEIRDLITGTOSSQASFIPCDIDT
            SOVYLPHELEKIRDLAENTHELMGNKIEIGWTECKITDNRKRPCHLVESKIPET
            EKNYNLDMSTETTLTLALGCHIAHNPAAEDLKKVYKPKYMSNGKRPAPDLSD
            VKLPDPOEVLDKLAENAHNVAKDRIKQWITGIDQDLNKNRNLRYVALDLDTK
            KSNDSLREAVRTFVGVYNIIPSDIDLAQDAVEKSIDKIFFEVMSVAVRSKWY
            FEEFVIVGDMRVGMARPGCRPIEELGADDOAEVPEGSGQMHGSGYFGRTMGD
            VEGCMINLDASMTFTLNGELITNKGSLAEADYIEENGVPICISLGSOTGRNGL
            MDASTFEKTYMGLOEGEPEPFAVNMNRDVAAMFSKRLPTFVNVPRDHPHIEVRIIDGT
            MDSPCLKVTHTKMGTONSNASMTVCRLSMPVPCSHSPCLDSEAROKRKOETI
            LSHHTTCYATIRFAGQDPSCVWYVTPDHYLSEKDLKNCITYVTYTLDEGVRV
            HESVRSNCTVMWGGDVIATSORSSRSNDLEIGCLVDLAMEHLSPSANGRLGTCYO
            VEPNTKVPVAFVLPQPTSTALEPFLKLNAMFLSAIKRSEKNVPOCPRLDVOT
            IOPLVMSRMSFPLKTEGVSERHGMVAVOCLEPLOMALIHIPERKRCVDLLECEOE
            DLMOFHHTIRIYSAVCALEGNHVAALGCHVDLSQFVAIDNKYLPGLIRSGFDL
            ISHLANAKERKLMMKNKNIITPITSATRKIRYDPSKSHGLDGLRCLRGPRFSL
            TPCVYVNGFEFHOKOSPEIRPLEIKTALSMLTEAVOGSHAHIRIDVQSGVFEQFPVY
            KLIGETLLVNGVPPDDDDIROLILLIDPSVCGHSEVETEDAEKEVYTOVEKAVENGK
            AKGAEPYKGLIOTRIPESVAKQKQCELLSTLDCEDLQREYATVAFGDTIVSLQINOK
            SRVIELMQALNMSAALFARKTRFERSPOEJINMLINLFQJGNCPCPEIRIELYDFH
            EDLLIHGCVPLDEEBEEDTSAWAKRLQVYKIKGPKRPEKQPEERKEPTPLKE
            LIISOTIRMAQOEDOIDAELVBMFSLRLQROYDSIGELLOALARKTYTISQASVPTIN
            LIALGQITRCLLSVAMKGEDELIMINGLOIMANNKVFYOHPMILVMEHLEVMVM
            DVLGAESQIAPPRKVAACCCFLCYCRISRONKAMPEHLSYLENSSVGLASMSR
            GSPPLDVASAVSVDNNELALGIEPRLVEVYTYIAGCGISGCPMLLACGYPRVGNPI
            EGERTYLSFLRPAFVNSSEVEENASVYKILIRPECTGPALRGEGEALAAAGAI
            KISENPALDPLSQGYKREVEDEDEIVHMGALMSFYSALIDLGRCAPMHLIOT
            GKEAIRIRSLIRSLVPTEDDVGIIISIPKLSLNDGSLISEPDMAANCCPDKAPMV
            LFLDRVIGIKDQFLLHLLEVGFLPDLRSASIDLVYSTSTPAALANVYLCSAVPL
            LTRCAPFAGTEHTYSLIDSTLOTIRLSKGRSLTAAODTIECLALCINLRMSL
            OOLIRLIPVDPOLNECYCKMLKILNHNHDECKYCYGLPSGMSGYLAAVEELHJTEK
            LFMGIPNLSHKKYVDPLFMALPCISATAGALPDPYIDPRTSATLEQISADQAGN
            DPKFIRTNMNSLPEKMEYITVTKAESHKRWACDSQNMKYGISLIDENVATPHILRP
            FKTIITEKETYRMPARESILTMALAVGSHVERKEDEALVQOREKRLKRSVQASQGN
            YSPAPLDLISNVVLSRELQGMVEVAENYHNIMAKKKLLELSKGGSHPLVYDYL

```

TAKEKFRDREKADOLEFKELVINGIVSRGVCMEDASSMEKFRAYKELIKLKYVDS
 AOEFLIAHLEAIVSSGKTEKSPHDOEIFEPAKVALPLVDVYFTNHILYFLSPSPKPLSS
 SGVASHKEKEMVASLFCKLALAYRHRISLFGSSTWVSCHLIAOTLDTPTVMSSG
 ELVKAGLRAPFENAAEDLEKTSBNLKLGTGHSRTQIKVSONINTVALLPILTSI
 FEHVAGHOFGVLLDLDVOISCHILCSYLSGTGKNIVYERORPALGCLSLAAEI
 PVAFLEPTLNRNPLSLFNTKPPRESISLIGMPTVEMCPDIPOLEGLKEKESDLAES
 GAVTEMPHIVETILMLCNYSYWMEREPREHLPSTGCMCTKYSHEHLSVILGNILK
 IINNNGIDIASWMKRILAVYAOPITISKARDILKSPHPLEKLAKKAYATVOEEOQL
 KADTKGDTQEAELLIDDEFVAVLCRDYAEPLMILRYVNNRSWMLKSPASDSTOERM
 VAEEVILMCKSHKREONFVIONETINNAFLTGDSKMSKSGODDERKTKRKG
 DLYSIOTSLTSLVAKKMLPLGLNMCTPGODELSLAKRSYSDTDEVEYEHRLNMLH
 LQKSDDPAYKQOLINTKYDLKSEPPNMEKTEVORVISAAPHLEVOYEDKTOYDLIP
 VMKLSQKQKRAVAVCFRPAAPLYNLPKHTNNEPILSTORVWLEKVENKTOYDLIP
 IIMKSPKEEEDDEEMKOPDPLHQITLHESNNALEKSLERLEDDPLITYSKSSMAKSC
 OSGEDEEEDDEKTEKEMERKOTLYOQARLHEGAALVOMISASGEGSPVAV
 ETLKGLAILNGNAGVOQKMLDYLEKKGAGFOSLSGIMOSCVLADLMAFEROKVA
 EGLMWTBEGTREKVLONDEFTDLFRFLOLLEGNHNSPONTLRONGMTTINVI
 STVDYLIRLOESISDFWYWSGDIIDESGONHESKALAVTKOIFNSLFEYIOGPCIG
 NOOSLARSRLMDAVYGLIHYAFANMOKLSODSSOIELKRLDLDLMDVMYMLSLLEG
 NVNNGTIGKQVDTLVSSSTNVEMILKFPFMPFLKIKDISSDPKEVDPDGKTIISK
 EFOKAMGQKQVDTLVSSSTNVEMILKFPFMPFLKIKDISSDPKEVDPDGKTIISK
 LSEHMPNDSRLKCLDPAGSVLNFEPYGLRIEMGAKRIEYVYELISSSTOMK
 POKESKROFIFDVNNGGEOEMKELFNCEDTIPEMOQASOISETDSKERDEE
 EDESSYVLEIEGEEDDKSFESASAFAMACASVKNVAFNLKATLKNLRKQRYAK
 KMYKELVYKFFSPFMMFLVGLPFLFTIIGLIPOLIMNTVFGGLVEGAKNIRVTKI
 LGMPPDPTGCIHNDALDAREAVTEPVGVPBELIHFKGKRGADIMSDGLFPPKKE
 GGLKHGPEGLGDLSEITIGKDEPTLSTYKRRKKRNOAEMKAKHEAGDVEPPKDL
 EDGEREDTAKEEDEQEAELMDVTKKRRRQKEREKPAFMANFKCLETIYQIKLHY
 LARNFYMLRFLALFAVAFINILFVYTEPEVEEEDTEVAVANLNSFTDEEBAE
 FVLESTGYMAPALRALAIHTVILSVGVYCLKVPALVFKREKEIARKELEGLY
 ITTEQSEDDIKGQDRVINTPSPNNWMDKFEVRYKINKYGDLYGAEIRLEELGLK
 NALDPSPEVASKAASIVSWLSIDMKYHMKGVYETDMSFLYLAHYTTMSVGLYK
 NNFFFAHLIDIANGFRTLRITISVTHNKOILVYGLIAYVAVITVVAENFERKE
 YNKSEDEDEPMKCDMMTCITLHMVYGVARGGIGTPEIDEPADPRXEMRYVNDITE
 FEFYVILLATIGLIIIDAREGLRQDOEVAREDETKCFITGINDNDPDTTPHCFETH
 TIOEHNLANYLFLFMYLILNKDETHGTQESYVMKYOERCDWDFPPADGCRKQYEDOL
 G"

BASE COUNT 3969 a 3667 c 3990 g 3442 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 4; Length 15068;
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGG 15
 DB 1316 ACCTCTTGATCGG 1302

RESULT 4
 CENS06JG2 1018 bp DNA linear SRS 10-JAN-2001
 LOCUS T7 end of clone AS0AA028E04 of library AS0AA from strain CLUB 533
 DEFINITION of Saccharomyces bayanus, sequence tagged site.
 AL401544
 AL401544.1 GI:12159319
 STS.
 Saccharomyces bayanus.
 Saccharomyces bayanus.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 1018)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolojin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Maupertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,M., Tekala,F., Toffano-Nioche,C., Weslowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL PERS Lett. 487 (1), 3-12 (2000)
 MEDLINE 20584711
 PUBMED 11152876

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

2 (bases 1 to 1018)
 Bon,E., Neuvéglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
 Aigle,M. and Durrens,P.
 Genomic exploration of the hemiascomycetous yeasts: 5.
 Saccharomyces bayanus var. uvarum
 PERS Lett. 487 (1), 37-41 (2000)
 20584715
 11152880
 3 (bases 1 to 1018)
 Genoscope.
 Direct Submission
 Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This STS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 Location/Qualifiers
 1..1018
 /organism="Saccharomyces bayanus"
 /strain="CLIB 533"
 /variety="uvarum"
 /db_xref="taxon:4931"
 /clone="AS0AA028E04"
 /clone_1ib="AS0AA"
 /note="end : 77"
 complement(<18..>422)
 /note="similar to Saccharomyces cerevisiae ORF YLL043w [Fps1 : glycerol channel protein]"
 /evidence=not-experimental
 misc_feature
 BASE COUNT 232 a 213 c 243 g 328 t 2 others
 ORIGIN

Query Match 93.3%; Score 14; DB 11; Length 1018;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGG 14
 DB 109 ACCTCTTGATCGG 122

RESULT 5
 E15997/c 1443 bp DNA linear PAT 28-JUL-1999
 LOCUS cDNA encoding human G protein-coupling receptor protein.
 DEFINITION E15997
 E15997.1 GI:5710680
 JF 1998127289-A/1.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1443)
 Hiyama,K., Fukuzumi,M. and Kawamata,Y.
 NEW G PROTEIN COUPLING TYPE RECEPTOR PROTEIN AND ITS DNA
 Patent: JP 1998127289-A 1 19-MAY-1998;
 TAKEDA CHEM IND LTD
 OS Homo sapiens (human)
 PN JP 1998127289-A/1
 PD 29-OCT-1998 JP 1996286823
 PF 29-OCT-1998 JP 1996286823
 PT HINDUA KUNJIT, FUKUZUMI MASASHI, KAWAMATA YUJI PC
 C12N15/09, C07H21/04, C07K14/705, C07K16/28, C12N1/21, C12P21/02, PC
 C12P21/08,
 PC C1201/68, G01N33/53, G01N33/566//A61K38/00, A61K39/395,

BASE COUNT	272 a	536 c	488 g	330 t
ORIGIN	/db_xref="taxon:9606"			
Query Match	93.3%;	Score 14;	DB 6;	Length 1626;
Best Local Similarity	100.0%;	Pred. No. 1.7e+03;		
Matches 14;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
Oy	2 CCTCTGGATCGGC 15			
Db	321 CCTCTGGATCGGC 308			
RESULT 7	HSEBTRLP2/c 1638 bp mRNA linear PRI 15-APR-1998			
LOCUS	HSEBTRLP2			
DEFINITION	Homo sapiens mRNA for G protein-coupled receptor ETBR-LP-2.			
ACCESSION	Y16280			
VERSION	Y16280.1 GI:3059117			
KEYWORDS	ETBR-LP-2 protein; G protein-coupled receptor.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
TITLE	1 (bases 1 to 1638)			
JOURNAL	Valdenaire,O., Giller,T., Breu,V., Ardail,A., Schweizer,A. and			
MEDLINE	Richards,J.G.			
PUBMED	A new family of orphan G protein-coupled receptors predominantly			
AUTHORS	expressed in the brain			
JOURNAL	FEBS Lett. 424 (3), 193-196 (1998)			
FEATURES	2 (bases 1 to 1638)			
ORIGIN	Valdenaire,O.			
DEFINITION	Submitted (22-JAN-1998) O. Valdenaire, F. Hoffmann-La Roche Ltd.,			
LOCUS	Prpv 69/34/8, grenzachstrasse, 124, CH-4070 Basel, SWITZERLAND			
FEATURES	Location/Qualifiers			
ORIGIN	1..1638			
DEFINITION	/organism="Homo sapiens"			
LOCUS	/db_xref="taxon:9606"			
DEFINITION	/clone="lambda GT10"			
LOCUS	52..1497			
DEFINITION	/function="G protein-coupled receptor"			
LOCUS	/codon_start=1			
DEFINITION	/product="ETBR-LP-2 protein"			
LOCUS	/protein_id="CA46153.1"			
DEFINITION	/db_xref="GI:3059118"			
LOCUS	/translation="MRMLPVLAVSLAVGLSRVSGAPLHLGRHRAETQEOOSRS			
DEFINITION	KGTDEPAKCVQGYVEENAAEYPRIPAGLOPRPLATSPNDKGGTDSGGL			
LOCUS	RNLGLGAGRLQIDNLELYPTSSSAVALMLALVYANIVANLSVMTIWNHAY			
DEFINITION	LKSANSLASIALMLADPLVLPFCPIPIVETITKQRLIDVSCRAVPMREYSIGYPT			
LOCUS	PSLCALGIDRHVATSTLPRVPRERKOSITLAKLAVIVGSGTTLAVPULLMOAOP			
DEFINITION	ATFTGTQDSCIMKPSASIPESYLSLVNNTONAMMMYRCGYCLDLETVTCQLVTKR			
LOCUS	VAGPGRKSECRASHKHCQESQLSNTYVGLTVVYACPLPENVNCIIVAAIVSTELTRQ			
DEFINITION	TLDLGLINQSTFEKGAITPVLLICICRPLDQAFIDCCCCCECGGASGASAAANG			
LOCUS	SNKTKTEVSSSTIFKHPRESPLPLPGTTC"			
BASE COUNT	273 a	526 c	493 g	346 t
ORIGIN				
Query Match	93.3%;	Score 14;	DB 9;	Length 1638;
Best Local Similarity	100.0%;	Pred. No. 1.7e+03;		
Matches 14;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
Oy	2 CCTCTGGATCGGC 15			
Db	189 CCTCTGGATCGGC 176			
RESULT 8	AX018426/c 2411 bp DNA linear PAT 07-SEP-2000			
LOCUS				

DEFINITION	Sequence 1 from Patent EP0943685.
ACCESSION	AX018426
VERSION	AX018426.1 GI:10042586
KEYWORDS	human.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 2411)
JOURNAL	Kroeger B.D. and Otterbach B.D.
	G-protein coupled receptor from human brain
	Patent: EP 0943685-A 1 22-SEP-1999;
	BasF AG (DE)
FEATURES	Location/Qualifiers
source	1..2411
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
5'UTR	1..19
CDS	20..1465
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAC07699.1"
	/db_xref="GI:10042589"
	/translation="MRMLPLAVSLAVTLAVGLSRVSGCAPLHGRHRAFTPOQSSK
	KRGDEDEAKGVQOYVPEENAEYRPRHPHGLOPTKPLVTSRPDKDSTPPSGQEL
	RGNLTGAPGQRLQIONLPYPTSSSAVAILMLALVFAVGLVGNLSVCKTVWSTY
	LKSNWSILSLALMDFLVLFCLPIVTFEITIKORLGLGVSCRAVFMVSSIGVTT
	FSCALGIDRFHVAFTSLPKPRPIEROSILAKLAVIWSMTLAVDELLMOIAQEP
	APRPGITDSCIMKPSASLPESLSLVMTYONARMMYFGCYFLPIFTVTCOLVTR
	VAPPGKSRSCRAKSHCOESOLNSTVGLVYAFCTLPENCNIVAVLSTELTRO
	TLDLGLINPSTFFPKATITPVLVLCICRPLGQAFDCCCCCECGGASLAAANG
	SDNMLKTEVSSSTFFHKRPRESPLPLPGTFC"
	1463..2411
BASE COUNT	397 a 775 c 659 g 580 t
ORIGIN	
3'UTR	
Query Match	93.3% Score 14; DB 6; length 2411;
Best Local Similarity	100.0%; Pred. NO. 1.7e+03;
Matches	14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	2 CCTCTGGATCGGC 15
Db	157 CCTCTGGATCGGC 144
RESULT 9	
AE001342	10585 bp DNA linear BCT 30-OCT-2000
LOCUS	Chlamydia trachomatis section 69 of the complete genome.
DEFINITION	AE001342 AE001273
ACCESSION	AE001342.1 GI:3329166
VERSION	
KEYWORDS	Chlamydia trachomatis.
SOURCE	Chlamydia trachomatis
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE	1 (bases 1 to 10585)
AUTHORS	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R.,
	Atyand L., Mitchell W.P., Olinger L., Tatsoy R.L., Zhao Q.,
	Koonin E.V. and Davis R.W.
TITLE	Genome sequence of an obligate intracellular pathogen of humans:
	Chlamydia trachomatis
JOURNAL	Science 282 (5389), 754-759 (1998)
MEDLINE	99000809
PUBMED	9784136
REFERENCE	2 (bases 1 to 10585)
AUTHORS	Kaiman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,
	Olinger L., Grimwood J., Davis R.W. and Stephens R.S.
TITLE	Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL	Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE	99206606
PUBMED	10192388
REFERENCE	3 (bases 1 to 10585)
AUTHORS	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R.,

TITLE	Direct Submission	Journal
Aravind, L., Mitchell, W.P., Olinger, L., Iatsov, R.L., Zhao, Y., Koonin, E.V. and Davis, R.W.	Submitted (20-May-1998)	Program in Infectious Diseases, University of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES	Location/Qualifiers	
source	1..10385	
gene	/organism="Chlamydia trachomatis"	
gene	/strain="D/UW-3/CX"	
gene	/db_xref="taxon:813"	
gene	88..2391	
gene	/gene="CT711"	
gene	88..2391	
gene	/gene="CT711"	
gene	/codon_start=1	
gene	/transl_table=1	
gene	/product="hypothetical protein"	
gene	/protein_id="AAC68306.1"	
gene	/db_xref="GI:3329167"	
gene	/translation="MSIQPSISLTKNITAFALAGEOVDAAYVMPQAVFPFOIDERS KGLKALGLEVDLEKFIPSLEKSPPIITGTSKISADGIEIYGLSELITLADPR KAAVQFEGGLADSFEDMLRLSENGGIDDPALAEETVIRYOTELNTLNRKIKOOSTL DDEYIKLAIPONVKFEKTESLKNNNNRNLIPKSRVNPQNMILTYNSVLSISEVPT GAMTAEVLSYLRATEFAAKLIREITNTKIDFNPVMDVREQTGIFGLKAEYNALIK NMIOGSLISFAGLFROLMSRPAVDEIGALYPRKDNEDYIHTAIDDYSLADLAKR EYVKLNGLSLTYAVAYSLGCFACKDQVFNMAQSLFDYTNPLNLOIOWMTPESTSES ISNVQALFTSKNPSADINDYGLTFDNKGLETNLENPFFPDYVSLMADPTKMSMODD TNPVIASESSIOKINQATTAWELELAIACGTRKAKLPPSLNPFNAVEAKKTFVTS IOMVYSMLDKYLPNOGYITLETLSQSMTFSNNAAYLNDIILAYAVSQTADAVYSLD LQVNSQGLDEPEVILSRANDYVKEKIDRSQADLFHCKAIEKIKELVSVNADYSLD SRALEETLASYAVEFENLYHNLSNYVWVSQVLSGVSKPDEVDEAFATKIGSEK OTRLOITLTFESAVTEGGRCNGVMGEGQVQLGLESKQDDYTSFNQOGLALQWESA IOQENTWMAALALMNQIFAKLIRRK"	
gene	2418..3590	
gene	/gene="CT712"	
gene	2418..3590	
gene	/gene="CT712"	
gene	/codon_start=1	
gene	/transl_table=1	
gene	/product="hypothetical protein"	
gene	/protein_id="AAC68307.1"	
gene	/db_xref="GI:3329168"	
gene	/translation="MRNHIPEGQKVFPTLETAAPERRRVSPAEEADYTLHEAAI YLVQFOLLNDANHOGLNKEEFVESLRDFTKTESMSLMQALMTESQREARRRKKI LOOLQESKYLQFOALTAEELHPVDDSLVNMKPEFQSAFVILDKYIIPAQEALRYAL RLTSQSYAQTLEFSPVLELTKSPNNAPPIVNLGSIIGQTAGTNFRFYGOVLDLRE ETGOLRKDIKNENAKQOLAOIKNIENANSLTEHKHTOLDKMANQIOTLDVGLISQ LNEFSTGLRGSLTFGRDETPATELMGSSSVYTLQNLBEKYVDGEINISSEPTGG LNEFTFLVADVQNFDDLOIOMLEOMRMAHQOQSLVYASLKLHNIRLASS"	
gene	complement(3616..4638)	
gene	/gene="porb"	
gene	/note="CT713"	
gene	complement(3616..4638)	
gene	/gene="porb"	
gene	/transl_table=1	
gene	/codon_start=1	
gene	/product="Outer Membrane Protein Analog"	
gene	/protein_id="AAC68308.1"	
gene	/db_xref="GI:3329169"	
gene	/translation="MSKRIYVILRLTFLSLGISTSLDAMPNGNPAPVINGINIE KNKSCFDLCNDSTOVLISALSGNKLKLCFCDDYIFSEEAQVRYKPYVTSYTAGNGSPSD TSTTKTRNFDLVNCLNNTNVCYAVAFSLPDRSLAIPLEFVSFEVKGGLQYTRLPMP AYDDFSEPLNASESEVTDGMIEVQNSGVFWDSLKYIWKDQSEFVGAGADRIASSQ PIDYIITANSQDVEFVLIADSDKLNFKEMSCVGLTTVYVDYVLYPLAISTGSYSRO PDSEFKLEDREPTNLKFKYRKITTSHRGNICGATNYVADNPFYNVEGRWGSGRAVN SGGFG"	
gene	complement(4771..5775)	
gene	/gene="gpda"	
gene	/note="CT714"	
gene	complement(4771..5775)	
gene	/gene="gpda"	
gene	/codon_start=1	
gene	/transl_table=1	

```

/product="Glycerol-3-P Dehydrogenase"
/protein_id="AAC68309.1"
/db_xref="GI:3329170"
/translation="MKETIATIGMMGFSEIANLANNGHVRVGNMARNPALTEGLSVQ
RRHRAAPHISIPONLSFTSHMEALDQATMVECVTSAGMRPVITQIKALTEURPVY
ITTSKIGRONTGLTSETALETEGRPAOHLYLSGPTASVLEKGCSCSVISAVND
TLKQIHRAFLTPFRVYVNSDLKGVALLGALKVIAIACGISDGFREGDNKSGLVNR
GIHEIRKPAITMGRPDTLNGIAGDICTSFSAFSRNTLFGKLAEGLETPEDAKK
IGMVEGYVYALSAHQIATNHRIDMPTTSVRYLVNLIQEGIAQLLRDRKEEYL"
complement(5772..7139)
/gene="CT715"
/complement(5772..7139)
/gene="CT715"
/codon_start=1
/transl_table=1
/product="AgX-1 Homolog-UDP-Glucose Pyrophosphorylase"
/protein_id="AAC68310.1"
/db_xref="GI:3329171"
/translation="MTDSYVDLSFLDQLPIQOEHLLEYPSLPOQRRLCTQIA
QIDIPFLQOALLQNPQASHQEVTPSPVYAGDNPAVYALQGLQLQKRGCVVIA
GGGSRILKFDGPKGLYPVSVKRPVQVLEKVAASKWRPLPLAINTSPLNHO
TLSYFATNDYFNLSQVDFPCDPLMLSLSGDLPLESEDRLSGITGNGCSTLLQ
SSGIWDKHOAGIEWVSYPIDNPLPDRDELGFPAHEENDVYITKTLROSOQEV
GVLLIEAKOKIAVVEYSYTLTKERCVKTEGDLTYKLANIGLYCLSDMLAQTAYOP
PLYKANKAKOLHSTTEPKNAKKEEPEFDLFQYSESAQVAVYRHRCEFAIKMYEEN
HSPATVREAMKREHALFTAVTERKLSPTNTELEADFPYPSHTSLEWETKTFQET
ITEAS"
complement(7151..7516)
/gene="CT716"
/complement(7151..7516)
/gene="CT716"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAC68311.1"
/db_xref="GI:3329172"
/translation="MNKRLQDLKSLTTELFKRRRTLETYKALSTIEHRLQOIQEH
AKISLTKHOKFLCRSTHEYDQHLQHLQREQTSLYKHQTLKTSLKDAYGDIQQLQD
RKIEKHDSYPIKSANN"
complement(7509..8813)
/gene="f111"
/note="CT717"
complement(7509..8813)
/gene="f111"
/codon_start=1
/transl_table=1
/product="Flagellum-specific ATP Synthase"
/protein_id="AAC68312.1"
/db_xref="GI:3329173"
/translation="MTHIQEFTLLHOMRPYREGCLSRISGSLTEAGLSACLEGIC
QISLSRDPILAIEVGINHRTLLALPIYUAIAGAVVULRPPASIPLSNHLGRY
LDGFNPIDGGPOLPKNTMSPLESPSPMSRPTIOEVPFGIAIDDLITIGGORY
GIFSEPGGKSSILSTAKGSOOTINVAITLGERREVDVNOHKEELARQRYVIA
SNAYETASKYVAGRAATTIAPEDGCAKPLTMDLSRMLSEIOEVAIKAGCTIS
HHIATSVFHVHAEFLERAGNDKSTISFYILHIANPDIPTDYVKLLDGHFELS
QEKSPSPINVLTSLSRSSQALPHYVAAQELSLKAVHEXIDIIYLGAVYSG
DAHIDRAIRLPSYKQFSOPYSYSAIHETIEOLCOLLKE"
complement(8887..9411)
/gene="CT718"
/complement(8887..9411)
/gene="CT718"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAC68313.1"
/db_xref="GI:3329174"
/translation="MEDYVAPSHRSLACINPOLPTPELOEALSHLSTIOVE
RKLVSHLSEIKESOKLPELLEAVVCEKFLYRKACPEELALISALOHILAT
YAVSPKIGLHEDLSNLSKWLITHDVPLKNIEFIADPCKKASYKIEPSSGIIIRD
IGELSHLSTVTP"
complement(9416..10420)
/gene="f11f"

```

```

CDS
/note="CT719"
complement(9416..10420)
/gene="f11f"
/codon_start=1
/transl_table=1
/product="Flagellar M-Ring Protein"
/protein_id="AAC68314.1"
/db_xref="GI:3329175"
/translation="MPVLPREFLKNIAVYKSLGYLLAAILIGFIMLYKPSPPQPVY
ASTEKPSHMLKSHLGNLSIEIQAKKEKDLTEFPYLOATVALSOEEDSLAEI
SVLSLPQASTLSPSLVHSITDYLRVSGTLKEHTLSDGNTLSPLFQSNLLF
TSLERISQTLTPOTHPALNIVADEGHOLLVDYDNTLTPKGVKLVLSMOEILS
APPEMPSVDYLPFLKPVAKHKSRLSSTLYLISGLIGCFAPYLAFTYDHS
QOREKIOSINIFRLIEMMRRESPEVALILSLDSAKABELLNKLPDEKSAVLKRT"
BASE COUNT 3114 a 1914 c 2513 g 3044 t
ORIGIN
Query Match 93.3%; Score 14; DB 1; Length 10585;
Best Local Similarity 100.0%; Pred No. 1.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTCTGGATCGG 14
Db 8395 ACCCTTGATCGG 8408
|||||
RESULT 10
MAC303116 16360 bp DNA circular MAM 23-JAN-2002
MAC303116
Tachygylossus aculeatus complete mitochondrial genome.
DEFINITION
ACCESSION
AJ303116
VERSION
AJ303116.1 GI:17932764
12S ribosomal RNA: 12S rRNA gene: 16S ribosomal RNA: 16S rRNA gene:
ATP synthase subunit 6: ATP synthase subunit 8: atp6 gene: atp8
gene: cot gene: co2 gene: cox3 gene: cytb gene: cytochrome b:
cytochrome c oxidase subunit 1: cytochrome c oxidase subunit 2:
cytochrome c oxidase subunit 3: internal transcribed spacer 1:
ITS1: NADH dehydrogenase subunit 1: NADH dehydrogenase subunit 2:
NADH dehydrogenase subunit 3: NADH dehydrogenase subunit 4: NADH
dehydrogenase subunit 4L: NADH dehydrogenase subunit 5: NADH
dehydrogenase subunit 6: NADH1 gene: NADH2 gene: NADH3 gene: NADH4
gene: NADH4L gene: NADH5 gene: NADH6 gene: transfer RNA-Ala:
transfer RNA-Ala gene: transfer RNA-Asn: transfer RNA-Asp: transfer
RNA-Cys: transfer RNA-Gln: transfer RNA-Glu: transfer RNA-Gly:
transfer RNA-His: transfer RNA-Ile: transfer RNA-Leu(CUN): transfer
RNA-Leu(UUR): transfer RNA-Lys: transfer RNA-Met: transfer RNA-Phe:
transfer RNA-Pro: transfer RNA-Ser(AGY): transfer RNA-Ser(UCN):
transfer RNA-Thr: transfer RNA-Trp: transfer RNA-Tyr: transfer
RNA-Val: tRNA-Ala gene: tRNA-Ala gene: tRNA-Asn gene: tRNA-Asp
gene: tRNA-Cys gene: tRNA-Gln gene: tRNA-Glu gene: tRNA-Gly gene:
tRNA-His gene: tRNA-Ile gene: tRNA-Leu(CUN) gene: tRNA-Leu(UUR)
gene: tRNA-Lys gene: tRNA-Met gene: tRNA-Phe gene: tRNA-Pro gene:
tRNA-Ser(AGY) gene: tRNA-Ser(UCN) gene: tRNA-Thr gene: tRNA-Trp
gene: tRNA-Tyr gene: tRNA-Val gene.
SOURCE
ORGANISM
Tachygylossus aculeatus
Mitochondrion Tachygylossus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Monotremata; Tachygylossidae; Tachygylossus.
REFERENCE
1
Janke,A., Magnell,O., Wiczorek,G., Westerman,M. and Arnason,U.
Phylogenetic Analysis of 18S rRNA and the Mitochondrial Genomes of
the Wombat, Vombatatus ursinus, and the Spiny Ant eater, Tachygylossus
aculeatus: Increased Support for the Marsupionta Hypothesis
J. Mol. Evol. 54 (1), 71-80 (2002)
JOURNAL MEDLINE PUBMED
21592585 11734900
2 (bases 1 to 16360)
REFERENCE
Janke, A.
Direct Submission
Submitted (21-DEC-2000) Janke A., Dept. of Genetics, Lund
University, Soelvegatan 29, S-22362 Lund, SWEDEN
LOCATION/Qualifiers
FEATURES

```

```

source      1. .16360
             /organism="Tachyglyssus aculeatus"
             /organelle="mitochondrion"
             /db_xref="taxon:9261"
             /tissue_type="liver"
             /dev_stage="adult"
             /country="Australia:Melbourne, Healesville Sanctuary"

gene        1. .72
             /gene="tRNA-Phe"

tRNA        1. .72
             /gene="tRNA-Phe"
             /product="tRNA-Phe"
             /note="codon recognized: UUC"
             /anticodon=(pos:35. .37,aa:Phe)

gene        73. .1022
             /gene="t2S rRNA"

rRNA        73. .1022
             /gene="t2S rRNA"
             /product="t2S rRNA"

gene        1080. .2646
             /gene="t6S rRNA"

rRNA        1080. .2646
             /gene="t6S rRNA"
             /product="t6S ribosomal RNA"

gene        2648. .2723
             /gene="tRNA-Val"
             /note="tRNA-Leu (UUR)"

tRNA        2648. .2723
             /gene="tRNA-Val"
             /product="tRNA-Val"
             /note="codon recognized: UUA"

misc_feature 2724. .2807
             /note="H-stand conserved spacer sequence, CSS"
             /function="unknown"

gene        2808. .3764
             /gene="nadh1"

CDS         2808. .3764
             /gene="nadh1"
             /codon_start=1
             /product="NADH dehydrogenase subunit 1"
             /protein_id="CAC88010.1"
             /db_xref="GI:17932765"

             /translation="MFLINLLILIVVLLAVAFLLIERKILIGYMOERKGPNIYCPHG
             LIDPIADAYKLTKEPRLPRTSSITYMFIAPILALSLATIIYPLMPPLIDINLGL
             LFLISVSGLSVYSILMSGWSNSKRYALTGALRAVAQTIISTEYFATILISIMLINSGF
             PLTTLNLTQEFMWLVVPTWPLMLMWFISTLAETNRAFPDITGESELSVGFNEYVAG
             PFAMFPLAEYANILIMNALTVILFEGYHILIFPELSTNFMVKIMLTSLFLMVRAS
             YPRRYDOLMHLMKNFLLPTLVTCIMYIMFPTMLSGTPPM"

gene        3765. .3833
             /gene="tRNA-Ile"

tRNA        3765. .3833
             /gene="tRNA-Ile"
             /product="tRNA-Ile"
             /note="codon recognized: AUC"
             /anticodon=(pos:3794. .3796,aa:Ile)

gene        complement(3831. .3902)
             /gene="tRNA-Gln"

tRNA        complement(3831. .3902)
             /gene="tRNA-Gln"
             /product="tRNA-Gln"
             /note="codon recognized: CAA"
             /anticodon=(pos:3867. .3869,aa:Gln)

gene        3902. .3970
             /gene="tRNA-Met"

tRNA        3902. .3970
             /gene="tRNA-Met"
             /product="tRNA-Met"
             /note="codon recognized: AUA"
             /anticodon=(pos:3923. .3925,aa:Met)

gene        3971. .5012
             /gene="nadh2"

CDS         3971. .5012
             /gene="nadh2"
             /note="TAA stop codon is completed by the addition of 3' A
             residues to the mRNA"
             /codon_start=1
             /translation="except=(pos:5012,aa:TERM)"
             /transl_table=2
             /product="NADH dehydrogenase subunit 2"
             /protein_id="CAC88011.1"
             /db_xref="GI:17932766"

             /translation="MTPMTNLIKMSSLKGMTITLSSHWLLMKCLEINTLAIPPL
             TSKRHPSTESAIRKFLTOATASMLMFLAASLNTWLGHWTLQDINNVVSMATRAL
             AMKGLAPFHYWYWEVLQGSPLSGMLITLQKLAIPSTIIYQISPLNNMTLLILAIS
             SILGNGNGLMOTOLRKIMAYSSIAHGNMIVIIITPPLTILNLVLYIMSTVALFW
             FYHINIKTKRPLSLMANKSPTMLMTTIVLISGLDPLTGFAPKMLVQDELIMHNI
             MMATTLAIMALNLFFYKRITYSSTLTTPPTNNKKHWKQPKKPLSLPPLTIST
             ALPLTPWFTYLG"

gene        5013. .5081
             /gene="tRNA-Trp"

tRNA        5013. .5081
             /gene="tRNA-Trp"
             /product="tRNA-Trp"
             /note="codon recognized: UCA"
             /anticodon=(pos:5043. .5045,aa:Trp)

gene        complement(5085. .5151)
             /gene="tRNA-Ala"

tRNA        complement(5085. .5151)
             /gene="tRNA-Ala"
             /product="tRNA-Ala"
             /note="codon recognized: GCA"
             /anticodon=(pos:5119. .5121,aa:Ala)

gene        complement(5152. .5524)
             /gene="tRNA-Asn"

tRNA        complement(5152. .5524)
             /gene="tRNA-Asn"
             /product="tRNA-Asn"
             /note="codon recognized: AAC"
             /anticodon=(pos:5189. .5191,aa:Asn)

gene        complement(5257. .5317)
             /gene="tRNA-Cys"

tRNA        complement(5257. .5317)
             /gene="tRNA-Cys"
             /product="tRNA-Cys"
             /note="codon recognized: UUU"
             /anticodon=(pos:5289. .5289,aa:Cys)

gene        complement(5318. .5383)
             /gene="tRNA-Tyr"

tRNA        complement(5318. .5383)
             /gene="tRNA-Tyr"
             /product="tRNA-Tyr"
             /note="codon recognized: UUU"
             /anticodon=(pos:5350. .5352,aa:Tyr)

gene        5384. .6925
             /gene="col1"

CDS         5384. .6925
             /gene="col1"
             /codon_start=1
             /transl_table=2
             /product="cytochrome c oxidase subunit 1"
             /protein_id="CAC88012.1"
             /db_xref="GI:17932767"

             /translation="MFINRWLFSTNKKDICTLYLLFGAWAGAGTALSLIRSELGQP
             GSLGSDQLYNVTVAHAFVMIFFKXMPIMIGFGFNGVLPLMIGAPDMAFPNNMMSF
             WLLPSPFLLLVSTVAGAGTGVVYPLACGLAHAGASVDLAIFSLHLACVSSITG
             AINFTITLINMKRPAMSOYOTPLFVWSVLTAVLLSLSPVLAAGITMLIDRNLTIT
             FPDAGGDBILYOHLEWFEFGHEEVYILILPGGILISHLVTVYSGKKEPFGYGMVWA
             MMSIGFLGFLVMAHMEVGMVDYTRAFTSAFTMLIAIPGVKVSMLATLHGDIKW
             TPRMLMAGIFLFTVGGVGLGIVLANSSDITLHDHYVVAHFHYVLSMGAVAFALMG
             FYWNPPLSGFTLHTTWAKVYVHTIMVGVNLTFFPDHFLGLAGMPRRYSDDPAYTLM
             NAISSGFSITLAVVWMTFVWEAFASREVLTVELTSTNEMLHGCPPTHTTEEP
             VYIKI"

gene        complement(6937. .6997)
             /gene="tRNA-Ser (UCN)"

```

trna	complement(6937..6997) /gene="trna-Ser (UCN)" /product="trna-Ser" /note="codon recognized: UCA" /anticodon=(pos:6963..6965,aa:Ser) 6997..7068 /gene="trna-Asp" 6997..7068 /gene="trna-Asp" /product="trna-Asp"				
gene					
trna					
Query Match	93.38; Score 14; DB 4; Length 16360; Best Local Similarity 100.0%; Pred. No. 1.6e+03; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	2 CCTCTGATCGGC 15 				
Db	9647 CCTCTGATCGGC 9660				
RESULT 11					
LOCUS	AP000389 55739 bp DNA linear PLN 27-DEC-2000				
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, p1 clone:MSJ3.				
ACCESSION	AP000389 BA000014				
VERSION	AP000389.1 GI:5672589				
KEYWORDS					
SOURCE	Arabidopsis thaliana (strain:Colombia) DNA, clone:lib:Mitsui P1 clone:MSJ3. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (sites)				
ORGANISM	Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,F. and Tabata,S. Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones DNA Res. 7 (3), 217-221 (2000)				
REFERENCE	2 (bases 1 to 55739) 20363099 Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,F. and Tabata,S. Direct Submission Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yakamuk@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MSJ3				
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT	Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-se (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-se/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is Mv11 and the 3' clone is M0P15. Location/Qualifiers 1..55739 /organism="Arabidopsis thaliana" /strain="Colombia" /db_xref="taxon:3702" /chromosome="3"				
FEATURES					
source					
CDS	/clone="MSJ3" /clone.lib="Mitsui P1" join(462..735,1345..1373) /note="gb AC61810.1 gene_id:MSJ3.1 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01420.1" /db_xref="GI:11994245" /translation="MDPNKARLSRLVLPDGLGEIVAKIGAKYAEVHNCILSKCEL GASANDERYLKLNLAPLVKKPLSKRKLILMKCLANNNDPAHITANSRNRK" complement(join(2420..4168,4328..5023,5138..5459)) /note="contains similarity to Arabidopsis thaliana retrotransposon Athila gene_id:MSJ3.2" /pseudo /codon_start=1 /evidence=not_experimental 6784..8961 /note="gene_id:MSJ3.3 unknown protein" /pseudo /codon_start=1 /evidence=not_experimental 8666..9013 /note="gb AAF24530.1 gene_id:MSJ3.4 similar to unknown protein" /pseudo /codon_start=1 /evidence=not_experimental join(13427..13562,13605..13769,14564..15219,15302..15409, 15500..15566,15681..15725,15808..16048,16127..16235, 16688..16859,16997..17124,17227..17382) /note="gb AAC97213.1 gene_id:MSJ3.5 strong similarity to unknown protein" /pseudo /codon_start=1 /evidence=not_experimental join(19136..20065,20144..20368,20467..20550,20807..20926, 21101..21220,21809..21955,22036..22182,22263..22353, 22432..22550,22624..22794,22894..22941) /note="gb AAD25853.1 gene_id:MSJ3.6 strong similarity to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01421.1" /db_xref="GI:11994246" /translation="MSTKKLFPRKLPFGSEVNDITFRVSKHSFKIVTVKALEPI EENTLMDTFLGIVFAEDLAFSSHIVHILORILTNNDLFEVADQPMRSLEB FITTGILPMDDSGSSPNVKKRKKKWIKNQTLKQKQDWMINKMPTLDLYYL EKRSKMDEERLRLVGLSVGLLANNPTTLPTELRKATPEKRCCKYPMNOYV SVLVNFRIGNDYLLRKOYGLFVQVSTQIMALSSYKLGKRGCTRESSGCMPII LQNRSTSSPSAKAITWADGDVHFLDSGFFVIOFKLPTTDFRRLIKDLIDVDVVI KQYRNSKKQWCGWIOVGHGRTDHLKPALETIVHQSRRKRSMSGKRRKRRK RKREDEERQEEEGKEELEVEYRGDGTKEQETPKGDEMEGEEKEEKEE BEKVEYRGEGTEKEQETPKGNEEMVEVEEKEEGKEEEOEKVLKEENVEEHD TSDQAVYVLSDEENGTAPTKEESOPKEETEVKREANYEHDHEETDOQAVYI LSDDNGAPAPERESPOKEETTEVPEETKDDVDVQVPLFMDEEETIGQSSLO PLTPVMLSEVWEETIDLVKKRKKKLLRLDLSIEELIMPDSKMOKEIVMDMDEK VLAALKRLLHLPKLLRAHREYVYVLEOIKTIMVEKDAATKAKNDLYV".25874, complement(join(23766..24009,24128..25699,25824..25874, 26025..26075,26157..26453)) /note="gene_id:MSJ3.7" /pseudo /codon_start=1 /evidence=not_experimental /product="Mutator-like transposase" 33238..33464 /note="contains similarity to Eu/Spm-like transposon				

protein
 gene_id:MSJ3.8"
 /pseudo
 /codon_start=1
 /evidence=not_experimental
 39943..40386
 /note="gene_id:MSJ3.9
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01422.1"
 /db_xref="GI:11994247"
 /translation="MAEFPHLPUNLTPEDPTEDEPADTLLIDDEVLYLHYIDSR
 SNHGVTPESTYDEIEGLVGNSSCTRRINAGEELPVYEPFTAEEMMERGLVYCALCRE
 ELVANEELSELPRHYHKECLSNLNSNTPCLCHNVELPNNH"
 42049..47107
 /note="gb|AAD15532.1
 gene_id:MSJ3.10
 strong similarity to unknown protein"
 /pseudo
 /codon_start=1
 /evidence=not_experimental
 complement(join(47610..47819,47908..48078,48155..48202))
 /note="gb|AAD23021.1
 gene_id:MSJ3.11
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01423.1"
 /db_xref="GI:11994248"
 /translation="MAAORSILFOGESSDDELDREIMRRTYGRNRRDRDKGVPIEDC
 CNKVVATSLDPTVEKLEPFSCPEISDGGCGGCKRMWVALCDERDMIKEETSE
 MKKDLAANKRVESDAEKIFMKKRETEKYESLNKYL"
 join(49962..50126,50196..50501)
 /note="gb|AAC69115.1
 gene_id:MSJ3.12
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01424.1"
 /db_xref="GI:11994249"
 /translation="MSSNSSENSIASTOSVSRMGRLPSKCHGCLDVIYITSTYAKKPG
 RPYRCPTRRDHLFRKVEDGTYEVVDSVPKLSVMSDELAKSEVAVELEMLQAMV
 KELEKRYRSEERINQIMIKLCCFLATIIIVIGIIMFPKINDEORFLGTY"
 complement(join(51037..51156,51256..51486))
 /note="contains similarity to En/Spm-like transposon
 protein
 gene_id:MSJ3.13"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01425.1"
 /db_xref="GI:11994250"
 /translation="WVPHIDGQFLRDVEVVLTPADRADLPVNLGLGTYDKTYTFW
 KFNHQLAVAGDILQRFNHAWPTLYOMIPVFKRMHTWDSRHLLTWTOFNLAARLY
 RDNMRALKRGRVY"
 join(52302..52436,52529..52793,52867..52969,53069..53312)
 /note="gene_id:MSJ3.14"
 /codon_start=1
 /evidence=not_experimental
 /product="replication protein A1-like"
 /protein_id="BAB01426.1"
 /db_xref="GI:11994251"
 /translation="MGYDGLRLDNPICITCWTICVMPFRMYISFTONSFLVAVFVDE
 WGIQIAIISRFQRYDMLLENOMKEIRTFGVYENGCVSRSTHKYIMELNATFV
 THSDPRETVLNPRTPEPDYIIEDTAHREINVDLGAIVDGCRLNITEDGEMDGFPL
 TPEIMDKYERKLCAGELAVDHRFRFRSFRNDSIYDINIFALITWKKLFRHYHTGLK
 HKVFONVAGISDVDPNIEVDELKAVLE"
 BASE COUNT 18008 a 10125 c 10679 g 16927 t
 ORIGIN

Query Match 93.3%; Score 14; DB 8; Length 55739;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTGGATCGGC 15
 Db 30106 CCTCTGGATCGGC 30093
 RESULT 12
 AC123621/C
 LOCUS
 DEFINITION
 AC123621
 VERSION
 AC123621.1 GI:21307021
 KEYWORDS
 HTG: HTGS. PHASE0.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 67079)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-74D13
 Unpublished
 2 (bases 1 to 67079)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Gold,S., Goyette,M., Graham,L.,
 Girard-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., MacDonald,P., Major,J., Marquis,N.,
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
 Meneus,L., Mihova,T., Mienga,Y., Murphy,T., Naylor,J., Nguyen,C.,
 Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkiang,P., Pierre,N.,
 Pollara,V., Raymond,C., Retta,R., Ribback,M., Riley,R., Rise,C.,
 Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S.,
 Schnapack,R., Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testafaye,S.,
 Theodore,J., Topham,K., Travers,M., Travis,N., Triggilio,J.,
 Vassiliev,B., Vlei,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zaioun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23301
 Center clone name: 74_D_13

NOTE: This record contains 85 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 698: contig of 698 bp in length
 * 699 798: gap of 100 bp
 * 799 1492: contig of 694 bp in length

* 1493 1592: gap of 100 bp
* 1593 2281: contig of 689 bp in length
* 2282 2381: gap of 100 bp
* 2382 3034: contig of 653 bp in length
* 3035 3134: gap of 100 bp
* 3135 3822: contig of 688 bp in length
* 3823 3922: gap of 100 bp
* 3923 4618: contig of 696 bp in length
* 4619 4718: gap of 100 bp
* 4719 5410: contig of 692 bp in length
* 5411 5510: gap of 100 bp
* 5511 6204: contig of 694 bp in length
* 6205 6304: gap of 100 bp
* 6305 6996: contig of 692 bp in length
* 6997 7096: gap of 100 bp
* 7097 7784: contig of 688 bp in length
* 7785 7884: gap of 100 bp
* 7885 8584: contig of 700 bp in length
* 8585 8684: gap of 100 bp
* 8685 9373: contig of 689 bp in length
* 9374 9473: gap of 100 bp
* 9474 10172: contig of 699 bp in length
* 10173 10272: gap of 100 bp
* 10273 10972: contig of 700 bp in length
* 10973 11072: gap of 100 bp
* 11073 11770: contig of 698 bp in length
* 11771 11870: gap of 100 bp
* 11871 12565: contig of 695 bp in length
* 12566 12665: gap of 100 bp
* 12666 13352: contig of 687 bp in length
* 13353 13452: gap of 100 bp
* 13453 14153: contig of 701 bp in length
* 14154 14253: gap of 100 bp
* 14254 14956: contig of 703 bp in length
* 14957 15056: gap of 100 bp
* 15057 15746: contig of 690 bp in length
* 15747 15846: gap of 100 bp
* 15847 16542: contig of 696 bp in length
* 16543 16642: gap of 100 bp
* 16643 17328: contig of 686 bp in length
* 17329 17428: gap of 100 bp
* 17429 18116: contig of 688 bp in length
* 18117 18216: gap of 100 bp
* 18217 18917: contig of 701 bp in length
* 18918 19017: gap of 100 bp
* 19018 19701: contig of 684 bp in length
* 19702 19801: gap of 100 bp
* 19802 20460: contig of 659 bp in length
* 20461 20560: gap of 100 bp
* 20561 21259: contig of 699 bp in length
* 21260 21359: gap of 100 bp
* 21360 22047: contig of 688 bp in length
* 22048 22147: gap of 100 bp
* 22148 22841: contig of 694 bp in length
* 22842 22941: gap of 100 bp
* 22942 23630: contig of 689 bp in length
* 23631 23730: gap of 100 bp
* 23731 24428: contig of 698 bp in length
* 24429 24528: gap of 100 bp
* 24529 25220: contig of 692 bp in length
* 25221 25320: gap of 100 bp
* 25321 26017: contig of 697 bp in length
* 26018 26117: gap of 100 bp
* 26118 26786: contig of 669 bp in length
* 26787 26886: gap of 100 bp
* 26887 27569: contig of 683 bp in length
* 27570 27669: gap of 100 bp
* 27670 28371: contig of 702 bp in length
* 28372 28471: gap of 100 bp
* 28472 29163: contig of 692 bp in length
* 29164 29263: gap of 100 bp
* 29264 29963: contig of 700 bp in length
* 29964 30063: gap of 100 bp

* 30064 30762: contig of 699 bp in length
* 30763 30862: gap of 100 bp
* 30863 31540: contig of 678 bp in length
* 31541 31640: gap of 100 bp
* 31641 32339: contig of 699 bp in length
* 32340 32439: gap of 100 bp
* 32440 33128: contig of 689 bp in length
* 33129 33228: gap of 100 bp
* 33229 33873: contig of 645 bp in length
* 33874 33973: gap of 100 bp
* 33974 34656: contig of 683 bp in length
* 34657 34756: gap of 100 bp
* 34757 35450: contig of 694 bp in length
* 35451 35550: gap of 100 bp
* 35551 36231: contig of 681 bp in length
* 36232 36331: gap of 100 bp
* 36332 37027: contig of 656 bp in length
* 37028 37127: gap of 100 bp
* 37128 37828: contig of 701 bp in length
* 37829 37928: gap of 100 bp
* 37929 38621: contig of 693 bp in length
* 38622 38721: gap of 100 bp
* 38722 39407: contig of 686 bp in length
* 39408 39507: gap of 100 bp
* 39508 40214: contig of 707 bp in length
* 40215 40314: gap of 100 bp
* 40315 41005: contig of 691 bp in length
* 41006 41105: gap of 100 bp
* 41106 41802: contig of 697 bp in length
* 41803 41902: gap of 100 bp
* 41903 42601: contig of 699 bp in length
* 42602 42701: gap of 100 bp
* 42702 43393: contig of 692 bp in length
* 43394 43493: gap of 100 bp
* 43494 44185: contig of 692 bp in length
* 44186 44285: gap of 100 bp
* 44286 44955: contig of 670 bp in length
* 44956 45055: gap of 100 bp
* 45056 45742: contig of 687 bp in length
* 45743 45842: gap of 100 bp
* 45843 46535: contig of 693 bp in length
* 46536 46635: gap of 100 bp
* 46636 47332: contig of 697 bp in length
* 47333 47432: gap of 100 bp
* 47433 48126: contig of 694 bp in length
* 48127 48226: gap of 100 bp
* 48227 48903: contig of 677 bp in length
* 48904 49003: gap of 100 bp
* 49004 49705: contig of 702 bp in length
* 49706 49805: gap of 100 bp
* 49806 50476: contig of 671 bp in length
* 50477 50576: gap of 100 bp
* 50577 51270: contig of 694 bp in length
* 51271 51370: gap of 100 bp
* 51371 52053: contig of 683 bp in length
* 52054 52153: gap of 100 bp
* 52154 52843: contig of 690 bp in length
* 52844 52943: gap of 100 bp
* 52944 53618: contig of 675 bp in length
* 53619 53718: gap of 100 bp
* 53719 54402: contig of 684 bp in length
* 54403 54502: gap of 100 bp

Query Match 93.3%: Score 14: DB 2: Length 67079;
Best Local Similarity 100.0%: Pred.No.1.5e+03;

Matches 14: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGC 15
|||||

Db 64842 CCTCTTGATCGGC 64829

RESULT 13

AL592300/c
LOCUS AL592300 68054 bp DNA linear PRI 28-FEB-2002
DEFINITION Human DNA sequence from clone RP11-294K24 on chromosome 1, complete sequence.
ACCESSION AL592300
KEYWORDS AL592300 GI:19068251
VERSION AL592300.16
SOURCE HTG.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 68054)
JOURNAL Direct Submission
Submitted (28-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquiries@sanger.ac.uk
humuqey@sanger.ac.uk
On Mar 1, 2002 this sequence version replaced gi:16973133.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-294K24 is from the library RP11-11.2 constructed by the group of Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-294K24. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-294K24 is at 1 in this sequence. The true left end of clone RP11-572A16 is at 66055 in this sequence.

FEATURES
source
1..68054
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-294K24"
/clone_11b="RP11-11.2"
BASE COUNT 14580 a 18489 c 18158 g 16827 t
ORIGIN

Query Match 93.3% Score 14: DB 9: Length 68054:
Best Local Similarity 100.0% Pred. No. 1.5e+03;
Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2 CCTCTTGATCGGC 15
|||||
Db 6184 CCTCTTGATCGGC 6171

RESULT 14
AC100880/c AC100880 69928 bp DNA linear HTG 23-NOV-2001
LOCUS AC100880
DEFINITION Mus musculus clone RP23-6716, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC100880
VERSION AC100880.1 GI:17059654
KEYWORDS HTG: HTGS_PPHASE0.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 69928)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
MUS musculus, clone RP23-6716
JOURNAL Unpublished
2 (bases 1 to 69928)
REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choquet, Y., Collingale, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Garayna, S., Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laboche, K., Lamazares, R., Landers, T., Lepocky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menzies, L., Milhova, T., Mlenka, V., Murphy, T., Nayler, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Polara, V., Raymond, C., Retica, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strausz, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainou, J., Zemke, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L14490
Center clone name: 67_1_6

NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
764 863: gap of 100 bp in length
* 864 1613: contig of 750 bp in length
* 1614 1713: gap of 100 bp
* 1714 2478: contig of 765 bp in length
* 2479 2578: gap of 100 bp
* 2579 3313: contig of 735 bp in length
* 3314 3413: gap of 100 bp
* 3414 4122: contig of 709 bp in length
* 4123 4222: gap of 100 bp
* 4223 4981: contig of 759 bp in length
* 4982 5081: gap of 100 bp
* 5082 5824: contig of 743 bp in length

```

* 5825 5924: gap of 100 bp
* 5925 6695: contig of 771 bp in length
* 6696 6795: gap of 100 bp
* 6796 7561: contig of 766 bp in length
* 7562 7661: gap of 100 bp
* 7662 8413: contig of 752 bp in length
* 8414 8513: gap of 100 bp
* 8514 9262: contig of 749 bp in length
* 9263 9362: gap of 100 bp
* 9363 10114: contig of 752 bp in length
* 10115 10214: gap of 100 bp
* 10215 10958: contig of 744 bp in length
* 10959 11058: gap of 100 bp
* 11059 11816: contig of 758 bp in length
* 11817 11916: gap of 100 bp
* 11917 12656: contig of 740 bp in length
* 12657 12756: gap of 100 bp
* 12757 13505: contig of 749 bp in length
* 13506 13605: gap of 100 bp
* 13606 14372: contig of 767 bp in length
* 14373 14472: gap of 100 bp
* 14473 15241: contig of 769 bp in length
* 15242 15341: gap of 100 bp
* 15342 16114: contig of 773 bp in length
* 16115 16214: gap of 100 bp
* 16215 16964: contig of 750 bp in length
* 16965 17064: gap of 100 bp
* 17065 17805: contig of 741 bp in length
* 17806 17905: gap of 100 bp
* 17906 18670: contig of 765 bp in length
* 18671 18770: gap of 100 bp
* 18771 19523: contig of 753 bp in length
* 19524 19623: gap of 100 bp
* 19624 20375: contig of 752 bp in length
* 20376 20475: gap of 100 bp
* 20476 21236: contig of 761 bp in length
* 21237 21336: gap of 100 bp
* 21337 22069: contig of 733 bp in length
* 22070 22169: gap of 100 bp
* 22170 22897: contig of 728 bp in length
* 22898 22997: gap of 100 bp
* 22998 23767: contig of 770 bp in length
* 23768 23867: gap of 100 bp
* 23868 24628: contig of 761 bp in length
* 24629 24728: gap of 100 bp
* 24729 25482: contig of 754 bp in length
* 25483 25582: gap of 100 bp
* 25583 26357: contig of 775 bp in length
* 26358 26457: gap of 100 bp
* 26458 27202: contig of 745 bp in length
* 27203 27302: gap of 100 bp
* 27303 28046: contig of 744 bp in length
* 28047 28146: gap of 100 bp
* 28147 28878: contig of 732 bp in length
* 28879 28978: gap of 100 bp
* 28979 29733: contig of 755 bp in length
* 29734 29833: gap of 100 bp
* 29834 30596: contig of 763 bp in length
* 30597 30696: gap of 100 bp
* 30697 31462: contig of 766 bp in length
* 31463 31562: gap of 100 bp
* 31563 32332: contig of 770 bp in length
* 32333 32432: gap of 100 bp
* 32433 33169: contig of 737 bp in length
* 33170 33269: gap of 100 bp
* 33270 34011: contig of 742 bp in length
* 34012 34111: gap of 100 bp
* 34112 34864: contig of 733 bp in length
* 34865 34964: gap of 100 bp
* 34965 35718: contig of 754 bp in length
* 35719 35818: gap of 100 bp
* 35819 36574: contig of 756 bp in length
* 36575 36674: gap of 100 bp

```

```

* 36675 37434: contig of 760 bp in length
* 37435 37534: gap of 100 bp
* 37535 38304: contig of 770 bp in length
* 38305 38404: gap of 100 bp
* 38405 39149: contig of 745 bp in length
* 39150 39249: gap of 100 bp
* 39250 39992: contig of 743 bp in length
* 39993 40092: gap of 100 bp
* 40093 40857: contig of 765 bp in length
* 40858 40957: gap of 100 bp
* 40958 41722: contig of 765 bp in length
* 41723 41822: gap of 100 bp
* 41823 42585: contig of 763 bp in length
* 42586 42685: gap of 100 bp
* 42686 43452: contig of 767 bp in length
* 43453 43552: gap of 100 bp
* 43553 44311: contig of 759 bp in length
* 44312 44411: gap of 100 bp
* 44412 45161: contig of 750 bp in length
* 45162 45261: gap of 100 bp
* 45262 46021: contig of 760 bp in length
* 46022 46121: gap of 100 bp
* 46122 46870: contig of 749 bp in length
* 46871 46970: gap of 100 bp
* 46971 47702: contig of 732 bp in length
* 47703 47802: gap of 100 bp
* 47803 48542: contig of 740 bp in length
* 48543 48642: gap of 100 bp
* 48643 49399: contig of 757 bp in length
* 49400 49499: gap of 100 bp
* 49500 50261: contig of 762 bp in length
* 50262 50361: gap of 100 bp
* 50362 51130: contig of 769 bp in length
* 51131 51230: gap of 100 bp
* 51231 51988: contig of 758 bp in length
* 51989 52088: gap of 100 bp
* 52089 52859: contig of 771 bp in length
* 52860 52959: gap of 100 bp
* 52960 53696: contig of 737 bp in length
* 53697 53796: gap of 100 bp
* 53797 54546: contig of 750 bp in length
* 54547 54646: gap of 100 bp
* 54647 55408: contig of 762 bp in length
* 55409 55508: gap of 100 bp
* 55509 56259: contig of 751 bp in length
* 56260 56359: gap of 100 bp
* 56360 57092: contig of 733 bp in length
* 57093 57192: gap of 100 bp
* 57193 57911: contig of 719 bp in length
* 57912 58011: gap of 100 bp
* 58012 58778: contig of 767 bp in length
* 58779 58878: gap of 100 bp

Query Match      93.3%: Score 14: DB 2: Length 69928;
Best Local Similarity 100.0%: Pred. No. 1.5e+03;
Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY      1 ACCTCTGATCGG 14
Db      27437 ACCTCTGATCGG 27424

RESULT 15
LOCUS   AL137862/c
DEFINITION Homo sapiens chromosome 11 clone RP5-1002E13 map p12-14.1, ***
ACCESSION AL137862
VERSION   AL137862.8 GI:16304424
KEYWORDS  HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE    human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


REFERENCE
AUTHORS
TITLE
JOURNAL

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 73509)

Walls, J.

Direct Submission

Submitted (16-Oct-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone

Requests: clonerequests@sanger.ac.uk

On Oct 21, 2001 this sequence version replaced gi:16214503.

COMMENT

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: dj1002E13

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 72684 bases at least Q40

Consensus quality: 73024 bases at least Q30

Insert size: 73509; sum-of-contrigs

Insert size: 93120; 6.4% error; agarose-fp

Quality coverage: 4.88x in Q20 bases; sum-of-contrigs Quality

coverage: 4.40x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

Location/Qualifiers

1..73509

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="p12-14.1"

/clone="RP5-1002E13"

/clone_1b="RPC1-5"

1..73509

/note="assembly_fragment:00544"

misc_feature 23859 a 13784 c 13831 g 22034 t 1 others

ORIGIN

Query Match 93.3%; Score 14; DB 2; Length 73509;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGG 14

DB 56207 ACCTCTTGATCGG 56194

Search completed: June 26, 2003, 03:47:27
Job time : 429.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds

(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-1

Perfect score: 1 accctctgcatgcgc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
N.Geneseq_101002: *
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	AA32387	Abl variable heavy
2	14	93.3	14	AA27174	C-protein coupled
3	14	93.3	14	AA30638	Human G protein co
4	14	93.3	14	AA30731	DNA encoding human
5	14	93.3	14	AA30731	Human EST-derived
6	14	93.3	14	AA30731	G-protein coupled
7	14	93.3	14	AA30731	Human muscleage cod
8	14	93.3	14	AA30731	CDNA encoding a G-
9	14	93.3	14	AA30731	Human brain G-prot

C 10	14	93.3	1038602	20	AA201425	Complete genome se
C 11	13.4	89.3	329	22	AA334099	Human CDNA encodin
C 12	13.4	89.3	360	15	AA066413	alca promoter. As
C 13	13.4	89.3	366	22	AA303486	DNA encoding anti-
C 14	13.4	89.3	375	22	AA303486	C glutamic codin
C 15	13.4	89.3	387	22	AA303486	DNA encoding anti-
C 16	13.4	89.3	390	24	AA303486	Human ovarian can
C 17	13.4	89.3	466	21	AA303486	Caenorhabditis ele
C 18	13.4	89.3	538	24	AA303486	Human benign prost
C 19	13.4	89.3	615	23	AA303486	Drosophila melanog
C 20	13.4	89.3	680	22	AA303486	Human CDNA clone (
C 21	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 22	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 23	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 24	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 25	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 26	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 27	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 28	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 29	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 30	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 31	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 32	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 33	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 34	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 35	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 36	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 37	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 38	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 39	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 40	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 41	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 42	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 43	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 44	13.4	89.3	819	21	AA303486	N. meningitidis pa
C 45	13.4	89.3	819	21	AA303486	N. meningitidis pa

ALIGNMENTS

RESULT 1	AA32387	Standard: DNA: 15 BP.
ID	AA32387	
XX	AA32387	
AC	AA32387	
XX	AA32387	
DT	17-JUN-1999	(first entry)
XX	17-JUN-1999	
DE	Abl variable heavy (VH) chain CDR1 encoding DNA.	
XX	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;	
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;	
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;	
KW	myelodysplasia; myelotoxic chemotherapy; leukemia; tumor; MDSK; CDR;	
KW	neuromuscular; muscular dystrophy; complementarity determining region;	
KW	variable heavy chain; variable light chain; VH; VL; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9910494-A2.	
XX		
PD	04-MAR-1999.	
XX		
PF	21-AUG-1998; 98MO-US17364.	
XX		
PR	25-AUG-1997; 97US-0918148.	
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;	
XX		
DR	WPI: 1999-204666/17.	

CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
 CC or Ala, and is preferably Lys. When the endogenous residue at this
 CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
 CC The 15 amino acid stretch between the substituted amino acid and the Pro
 CC may be endogenous, non-endogenous, or a mixture of endogenous and
 CC non-endogenous residues. The constitutively active GPCRs are useful for
 CC identifying antagonists, agonists and partial agonists for use as
 CC pharmaceutical agents. The mutant proteins are also useful in research
 CC settings for elucidating the roles of the receptors in normal and
 CC diseased conditions. Antagonists for a particular GPCR are useful for
 CC treating diseases and disorders associated with that receptor. Because
 CC the novel mutant GPCRs are constitutively active, they can be used
 CC directly for screening of compounds without the need for endogenous
 CC ligands. The present sequence represents cDNA encoding a human wild-type
 CC GPCR used in an exemplification of the invention. This was cloned and
 CC subjected to site-directed mutagenesis (SDM) to generate DNA encoding
 CC the corresponding mutant of the invention.

XX SQ Sequence 1446 BP; 250 A; 473 C; 426 G; 297 T; 0 other;

Query Match 93.3%; Score 14; DB 21; Length 1446;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

YY 2 CCTCTTGATCGGC 15
 |||
 DB 138 CCTCTTGATCGGC 125

RESULT 4
 AAA30731/C
 ID AAA30731 standard; DNA; 1446 BP.

AC AAA30731;

DT 21-AUG-2000 (first entry)

XX DNA encoding human mutant G protein-coupled receptor ETRB-LP2 (N358K).

DE G protein-coupled receptor; GPCR; constitutively active;

KM intracellular loop 3; transmembrane domain 6; drug screening;

XX agonist; antagonist; mutant; ss.

OS Homo sapiens.

XX Synthetic.

XX WO200022129-A1.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-US23938.

XX 13-OCT-1998; 9805-0170496.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw CW;

XX WPI; 2000-329165/28.

XX P-PSDB; AAY90665.

XX Non-endogenous constitutively activated human G protein-coupled

XX receptors, useful for identifying agonists for use as pharmaceutical

XX agents

XX Example 2; Page 263-264; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions

XX of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-

XX AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743

XX and AAA30775-A30779). The mutant proteins of the invention contain a

XX mutation in a portion of the protein comprising intracellular loop 3

XX (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,

CC is substituted for an endogenous residue in IC3 at a position 16 amino
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence
 CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
 CC or Ala, and is preferably Lys. When the endogenous residue at this
 CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
 CC The 15 amino acid stretch between the substituted amino acid and the Pro
 CC may be endogenous, non-endogenous, or a mixture of endogenous and
 CC non-endogenous residues. The constitutively active GPCRs are useful for
 CC identifying antagonists, agonists and partial agonists for use as
 CC pharmaceutical agents. The mutant proteins are also useful in research
 CC settings for elucidating the roles of the receptors in normal and
 CC diseased conditions. Antagonists for a particular GPCR are useful for
 CC treating diseases and disorders associated with that receptor. Because
 CC the novel mutant GPCRs are constitutively active, they can be used
 CC directly for screening of compounds without the need for endogenous
 CC ligands. Sequences AAA30709-AAA30743 and AAA30775-A30779 represent DNAs
 CC encoding the mutant human GPCRs of the invention.

XX SQ Sequence 1446 BP; 250 A; 472 C; 427 G; 297 T; 0 other;

Query Match 93.3%; Score 14; DB 21; Length 1446;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

YY 2 CCTCTTGATCGGC 15
 |||
 DB 138 CCTCTTGATCGGC 125

RESULT 5
 AAH98401/C
 ID AAH98401 standard; cDNA; 1616 BP.

AC AAH98401;

DT 12-OCT-2001 (first entry)

XX Human EST-derived coding sequence SEQ ID NO: 258.

DE Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KM tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

XX diagnostics; forensic test; gene mapping; genetic disorder;

XX biodiversity; gene therapy; nutrition; ss.

XX Homo sapiens.

XX WO200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02687.

XX 25-JAN-2000; 2000US-0491404.

XX 17-JUL-2000; 2000US-0617746.

XX 03-AUG-2000; 2000US-0631451.

XX 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.

XX P-PSDB; AAM23742.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising

XX antibodies and research use

XX Claim 1; Page 368-369; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel

XX proteins from a variety of organisms, including human, dog, cat, horse,

XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.

SO Sequence 1616 BP; 269 A; 526 C; 480 G; 341 T; 0 other;

Query Match

Best Local Similarity 93.3%; Score 14; DB 22; Length 1616;
 Pred. No. 2.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 DB 166 CCTCTTGATCGGC 153

RESULT 6

AAV27175/c
 ID AAV27175 standard; cDNA: 1626 BP.

XX AAV27175;

DT 17-SEP-1998 (first entry)

DE G-protein coupled receptor, long form, coding sequence.

KW G-protein coupled receptor; gene therapy; abnormality detection;
 KW long form; human; ds.

OS Homo sapiens.

PN EP845529-A2.

XX 03-JUN-1998.

PF 27-OCT-1997; 97EP-0308562.

PR 29-OCT-1996; 96JP-0286823.

XX (TAKE) TAKEDA CHEM IND LTD.

PI Fukusumi S, Hinuma S, Kawamata Y;

DR WPI: 1998-288746/26.

P-PSDB: AAM55030.

PT New human G-protein coupled receptor protein - and corresponding
 PT DNA, ligands, antibodies, etc

PS Claim 6; Page 50-51; 65pp; English.

CC This sequence encodes a human G-protein coupled receptor of the
 CC invention. The protein or cells expressing the DNA encoding it can be
 CC used to screen for agonists or antagonists of the receptor, which can be
 CC also be used for treating various diseases (none disclosed). The DNA can
 CC also be used for practice drug design based on comparisons with
 CC structurally analogous ligands and receptors. DNA encoding the protein
 CC can be used for gene therapy for diseases caused by a deficiency of the
 CC receptor. The DNA can also be used to detect abnormalities in the gene
 CC encoding the receptor. The protein or fragment can be used to determine
 CC levels of receptor ligands in vivo. The antibody can be used in assays to
 CC detect the protein.

SO Sequence 1626 BP; 272 A; 540 C; 484 G; 330 T; 0 other;

Query Match

Best Local Similarity 93.3%; Score 14; DB 19; Length 1626;
 Pred. No. 2.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 DB 321 CCTCTTGATCGGC 308

RESULT 7

AAZ11093/c
 ID AAZ11093 standard; cDNA: 1728 BP.

XX AAZ11093;

DT 03-NOV-1999 (first entry)

DE Human mucilage coding sequence.

KW Mucilage; human; G-protein-coupled receptor; infection; cancer; diabetes;
 KW obesity; anorexia; Parkinson's disease; hypotension; hypertension;
 KW osteoporosis; stroke; myocardial infarction; asthma; allergy; therapy;
 KW neurological disorder; dyskinesia; diagnosis; ss.

XX Homo sapiens.

OS

XX Key

XX CDS

FT Location/Qualifiers
 FT 150..1595
 FT /*tag= a
 FT /product= mucilage

PN W09942485-A1.

PD 26-AUG-1999.

PF 17-FEB-1999; 99MO-US03320.

PR 08-MAY-1998; 98US-0075468.

PR 20-FEB-1998; 98US-0075307.

XX (SMIK) SMITHKLINE BEECHAM CORP.

PI Elshourbagy N, Vawter L;

DR WPI: 1999-527457/44.

P-PSDB: AAY32886.

PT New mucilage receptor polypeptide for treatment, prevention and
 PT diagnosis of e.g. infections and cancer

XX Claim 2; Page 34; 50pp; English.

CC This sequence encodes the mucilage protein of the invention. Mucilage is
 CC a G-protein-coupled receptor. Mucilage, or the nucleic acid encoding it,
 CC are used to treat conditions requiring increased activity/expression of
 CC mucilage, while antagonists (or nucleic acid encoding them or competitive
 CC peptides) of the protein are used where the opposite effect is required.
 CC Typical of very many diseases that can be treated are infections (viral,
 CC bacterial, fungal, protozoal, particularly human immune deficiency
 CC virus-1 or HIV-2); pain; cancer; diabetes; obesity; anorexia; Parkinson's
 CC disease; hypotension or hypertension; osteoporosis; stroke; myocardial
 CC infarction; asthma; allergy; neurological disorders; dyskinesia etc.
 CC Detecting mutations in the DNA and/or determining the presence or amount
 CC of the protein is used to diagnose these diseases, or susceptibility to
 CC them. The mucilage protein is also used to screen for specific
 CC (antagonists, which are potential therapeutic agents; to raise
 CC antibodies (Ab); in protective vaccines; to isolate cognate receptors;
 CC and in structure-based drug design. The mucilage coding sequences are
 CC used to produce recombinant mucilage; as a source of primers and probes
 CC for isolating related genes or detecting mutations; for chromosome
 CC identification; and as source of antagonistic oligonucleotides. The Ab
 CC are used to identify or isolate mucilage-expressing cells, for affinity
 CC purification of mucilage and therapeutically.

SO Sequence 1728 BP; 284 A; 569 C; 513 G; 362 T; 0 other;

Query Match

Best Local Similarity 93.3%; Score 14; DB 20; Length 1728;
 Pred. No. 2.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 DB 287 CCTCTTGATCGGC 274

RESULT 8
 AA236408/C
 ID AA236408 standard: cDNA; 1850 BP.

AC AA236408;

XX 22-FEB-2000 (first entry)

DE cDNA encoding a G-protein coupled receptor designated HG01.

KW Human; G-protein coupled receptor; HG01; endothelin receptor;
 KW pathological state; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS /tag= a
 FT 84..1529 /product= "G-protein coupled receptor HG01"

FT sig_peptide /transl_except= (pos: 186..188, aa: Ala)

FT mat_peptide /tag= b
 FT 159..1526 /product= "G-protein coupled receptor"

FT /tag= c /note= "specifically claimed in claim 2"

XX WO960153-A2.

XX 25-NOV-1999.

XX 17-MAY-1999; 99WO-US10808.

XX 21-MAY-1998; 98US-0086294.

XX (MERI) MERCK & CO INC.

XX Liu Q;

XX WPI: 2000-053303/04.

XX P-PSDB: AAY53779.

XX Novel receptor protein, useful for identifying antagonists and agonists
 for drug development

XX Claim 2: Fig 1A; 33pp; English.

XX The present sequence encodes a human G-protein coupled receptor,
 designated HG01. HG01 is highly expressed in the brain, and is thought
 to be related to the endothelin receptors. The HG01 cDNA sequence was
 isolated from a human foetal brain library. HG01 can be used to
 identify binding substances, antagonists and agonists, which may be
 useful pharmacologically to modulate HG01 activity. G-protein coupled
 receptors (GPCRs) are involved in many pathological states, and
 modulators of GPCR activity are therefore useful to treat such
 conditions. HG01 (or fragments) are also useful to produce antibodies.
 CC The polynucleotides can be used in gene therapy, especially to treat
 CC conditions requiring increased HG01.

XX Sequence 1850 BP; 305 A; 614 C; 540 G; 390 T; 1 other;

XX Query Match 93.3%; Score 14; DB 21; Length 1850;

XX Best Local Similarity 100.0%; Pred. NO. 2.3e+02;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 DB 221 CCTCTTGATCGGC 208

RESULT 9
 AA200860/C
 ID AA200860 standard: cDNA to mRNA; 2411 BP.

AC AA200860;

XX 11-OCT-1999 (first entry)

DE Human brain G-protein coupled receptor cDNA.

KW G-protein coupled receptor; human; brain; anti-neurodegeneration; pain;
 KW antitumor; analgesic; anti-epileptic; anti-addictive; anti-obesity;
 KW anti-anorexic; anti-bulimic; cerebroprotective; gene therapy; diagnostic;
 KW detection; disease; central nervous system; peripheral nervous system;
 KW Alzheimer's disease; Parkinson's disease; brain tumour; schizophrenia;
 KW epilepsy; drug addiction; obesity; anorexia; bulimia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..19
 FT /tag= a

FT CDS 20..1465 /tag= b
 FT /product= "G-protein coupled receptor"

FT 3'UTR 1466..2411 /tag= c

XX DE19805351-A1.

XX 12-AUG-1999.

XX 11-FEB-1998; 98DE-1005351.

XX 11-FEB-1998; 98DE-1005351.

XX (BAD) BASF AG.

XX Kroegeer B, Otterbach B;

XX WPI: 1999-445442/38.

XX P-PSDB: AAY25969.

XX New human G-protein coupled receptor from brain tissue, used to
 treat nervous system disorders, e.g. Alzheimer's disease, eating
 disorders and as cerebral protectant

XX Claim 5: Page 6-8; 12pp; German.

XX This invention describes a novel G-protein coupled receptor isolated
 from human brain tissue which has anti-neurodegeneration, antitumor,
 analgesic, anti-epileptic, anti-addictive, anti-obesity, anti-anorexic,
 anti-bulimic and cerebroprotective activity. Nucleic acid (I) that
 encodes (I), or its complement, is used (1) in gene therapy of
 (1)-deficiency states, (2) as probes and primers for diagnostic
 detection of (I) in usual hybridization/amplification assays, (3) for
 recombinant expression of (I), (4) to isolate the corresponding genomic
 sequence, (5) to design ribozymes and antisense sequences for treating
 overexpression of (I), (6) to produce transgenic animals for studying
 the (patho)physiology of (I) and (7) to detect mutations in (I)-encoding
 genes. (I) is used (1) to raise specific antibodies (Ab), (2) to screen
 for specific ligands, agonists and antagonists, (potentially useful as
 therapeutic agents) and (3) to treat (1)-deficiency conditions. Abs to
 (I) are used to detect/determine (I) in standard immunoassays,
 CC particularly for diagnosis. (I) is associated with chronic and acute
 diseases of the central and peripheral nervous systems (e.g. Alzheimer's
 or Parkinson's diseases, brain tumours, pain, schizophrenia, epilepsy,
 CC drug addiction etc.), diseases that include a nervous component (e.g.
 CC obesity, anorexia and bulimia). This sequence encodes the human brain
 CC G-protein coupled receptor described in the invention.

XX Sequence 2411 BP; 397 A; 775 C; 659 G; 580 T; 0 other;

Query Match 93.3%; Score 14; DB 20; Length 2411;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 CCTCTTGATCGGC 15
157 CCTCTTGATCGGC 144

RESULT 10
AAZ01425/c
ID AAZ01425 standard; DNA: 1038602 BP.

XX AC AAZ01425;
XX DT 07-OCT-1999 (first entry)
XX DE Complete genome sequence of Chlamydia trachomatis.
XX KM Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX KM paratrachoma; inclusion conjunctivitis; genital disease; peritrititis;
XX KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX KM bartolinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX OS Chlamydia trachomatis.
XX PN WO9928475-A2.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98MO-1B01939.
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97FR-0015041.
XX PR 17-DEC-1997; 97FR-0016034.
XX PA (GEST) GENSET.
XX PI Griffais R;
XX DR WPI; 1999-371125/31.
XX PT Genome sequence of Chlamydia trachomatis
XX PS Claim 1; Page 373-656; 1755pp; English.
XX CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conjunctivitis; genital diseases such as nongonococcal urethritis,
CC epididymitis, cervicitis, salpingitis, peritrititis, bartolinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 93.3%; Score 14; DB 20; Length 1038602;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACCTCTTGATCGG 14
683576 ACCTCTTGATCGG 683563

RESULT 11
AAS34099/c

ID AAS34099 standard; cDNA; 329 BP.
XX AAS34099;
XX AC AAS34099;
XX DT 17-DEC-2001 (first entry)
XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 623.
XX KM Human: foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KM immunomodulator; cardiovascular; cytostatic; nephrothropic;
KM cardiovascular; autoimmune disease; rheumatoid arthritis;
KM hyperproliferative disorder; breast neoplasm; cancer;
KM cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KM cerebral ischaemia; angiogenesis; nervous system disorder;
KM Alzheimer's disease; infection; ocular disorder; corneal infection;
KM wound healing; epithelial cell proliferation; food additive.
XX OS Homo sapiens.
XX PN WO200155312-A2.
XX PD 02-AUG-2001.
XX PE 17-JAN-2001; 2001MO-US01321.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214866.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226271.
XX PR 22-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226868.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 01-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231245.
XX PR 08-SEP-2000; 2000US-0231413.


```

XX Promoter; alca; crea binding site; glucose repression; ds.
KM Aspergillus nidulans.
OS
XX WO9413820-A.
XX
XX 23-JUN-1994.
XX
XX 10-DEC-1993; 93WO-EP03553.
XX
XX 10-DEC-1992; 92US-0988778.
XX
XX (KONN ) GIST-BROCADES NV.
XX
XX Hintz WE, Lagosky PA;
XX
XX WPI; 1994-217893/26.
XX
XX DNA constructs comprising a filamentous fungal promoter variant
PT having a disrupted crea binding site - mediates expression of the
PT protein in the presence of glucose
XX
XX PS Disclosure; Page 13; 21pp; English.
XX
XX A region of the A. nidulans alca promoter is given in AA066413. To
CC construct an alca promoter variant, the crea binding sites were
CC disrupted, but spatially conserved, by replacement of residues.
CC The promoter variant allows expression of foreign genes in
CC Aspergillus nidulans without the need to establish glucose-depleted
CC growth conditions.
XX
XX Sequence 360 BP; 97 A; 111 C; 83 G; 69 T; 0 other:
SO
Query Match 89.3%; Score 13.4; DB 15; Length 360;
Best Local Similarity 93.3%; Pred. No. 4.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ACCTCTGGATCGGC 15
Db 106 ACCTCTGGATCGGC 92
RESULT 13
AA503486
ID AA503486 standard; CDNA; 366 BP.
XX
XX AA503486;
AC
XX
XX 29-AUG-2001 (first entry)
DT
XX
XX DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 73.
DE
XX
XX Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
KM heart disease; complementarity determining region; CDR; ss.
XX
XX Homo sapiens.
OS
XX WO200127279-A1.
XX
XX 19-APR-2001.
XX
XX 11-OCT-2000; 2000WO-GB03900.
XX
XX 12-OCT-1999; 99US-0158812.
XX
XX (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
PA
XX Edwards BM, Main SH, Vaughan TJ;
PI
XX WPI; 2001-282031/29.
XX
XX P-PSDB; AAU02586.
XX

```

```

PT Panel of specific binding members of antibody molecules which bind to
PT whole adipocytes is used in the treatment of obesity and obesity
PT related diseases -
XX
XX PS Disclosure; Page 146; 182pp; English.
XX
XX AA503401-AA503535 represent anti-adipocyte monoclonal antibody heavy
CC and light chain coding sequences of the invention. The antibodies can be
CC used in the treatment of obesity and obesity related diseases. The
CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
CC mass of an obese patient or the antibody can be used as a therapeutic
CC itself. Antibodies binding specifically to adipocytes can be used to
CC activate the immune system to destroy the cells by complement mediated
CC lysis. The antibodies may be labeled with a detectable label such as
CC radiolabel, fluorescent or chemical group and used in methods of
CC diagnosis in human subjects e.g. to determine the presence of adipocyte
CC antigen on the surface of an adipocyte to detect or determine the
CC presence or level of adipocytes in a cell or tissue sample. The
CC antibodies can be used as an alternative means of treatment for obese
CC patients other than undergoing surgery to remove excess fat. Antibodies
CC for different types of fat deposits can also be produced e.g. intra-
CC abdominal fat associated with heart disease.
XX
XX Sequence 366 BP; 89 A; 97 C; 101 G; 79 T; 0 other:
SO
Query Match 89.3%; Score 13.4; DB 22; Length 366;
Best Local Similarity 93.3%; Pred. No. 4.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ACCTCTGGATCGGC 15
Db 91 ACCTATTGGATCGGC 105
RESULT 14
AAH66453
ID AAH66453 standard; DNA; 375 BP.
XX
XX AAH66453;
AC
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 1488.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (RYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX
XX P-PSDB; AAG91234.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 8; SEQ ID NO: 1488; 246pp + Sequence Listing; English.
XX

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-1
Perfect score: 15
Sequence: 1 acctcttgatcgcc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	US-08-918-148-1	Sequence 1, Appli
2	14	93.3	1443	US-08-959-381A-3	Sequence 3, Appli
3	14	93.3	1626	US-08-959-381A-4	Sequence 4, Appli
4	13.4	89.3	39	US-08-448-619-2	Sequence 2, Appli
5	13.4	89.3	39	US-08-448-619-3	Sequence 3, Appli
6	13.4	89.3	60	US-08-470-058-11	Sequence 11, Appl
7	13.4	89.3	360	US-08-321-474-10	Sequence 10, Appl
8	13.4	89.3	1177	US-08-841-349-11	Sequence 11, Appl
9	13.4	89.3	4403765	US-09-103-840A-2	Sequence 10, Appl
10	13.4	89.3	4411529	US-09-103-840A-1	Sequence 1, Appli
11	13	86.7	9919	US-08-880-179-1	Sequence 1, Appli
12	13	86.7	9936	US-08-972-927-2	Sequence 2, Appli
13	12.4	82.7	213	US-08-247-907A-3	Sequence 3, Appli
14	12.4	82.7	213	US-08-452-772-3	Sequence 3, Appli
15	12.4	82.7	213	US-09-414-234-3	Sequence 3, Appli
16	12.4	82.7	213	US-08-919-850-3	Sequence 3, Appli
17	12.4	82.7	213	PCT-US94-05288-3	Sequence 3, Appli
18	12.4	82.7	388	US-09-056-556-178	Sequence 178, App
19	12.4	82.7	388	US-09-072-566-173	Sequence 173, App
20	12.4	82.7	445	US-09-385-962-467	Sequence 467, App
21	12.4	82.7	472	US-08-975-316-16	Sequence 16, Appl
22	12.4	82.7	472	US-09-615-192A-16	Sequence 16, Appl
23	12.4	82.7	567	US-09-615-192A-16	Sequence 100, App
24	12.4	82.7	917	US-08-474-633A-6	Sequence 6, Appli
25	12.4	82.7	917	US-08-823-771-6	Sequence 6, Appli
26	12.4	82.7	917	PCT-US93-02480-6	Sequence 6, Appli
27	12.4	82.7	985	US-09-056-556-182	Sequence 182, App

28	12.4	82.7	985	US-09-072-596-177	Sequence 177, App
29	12.4	82.7	1230	US-09-387-574-5	Sequence 5, Appli
30	12.4	82.7	1230	US-09-668-096-5	Sequence 5, Appli
31	12.4	82.7	1270	US-08-247-907A-10	Sequence 10, Appl
32	12.4	82.7	1270	US-08-452-772-10	Sequence 10, Appl
33	12.4	82.7	1270	US-09-414-234-10	Sequence 10, Appl
34	12.4	82.7	1270	US-08-919-850-10	Sequence 10, Appl
35	12.4	82.7	1270	PCT-US94-05288-10	Sequence 10, Appl
36	12.4	82.7	1393	US-08-765-875-1	Sequence 1, Appli
37	12.4	82.7	1393	US-08-795-671-1	Sequence 1, Appli
38	12.4	82.7	1411	US-08-674-168-18	Sequence 18, Appl
39	12.4	82.7	1411	US-08-985-908-14	Sequence 14, Appl
40	12.4	82.7	1411	US-08-852-730-19	Sequence 19, Appl
41	12.4	82.7	1455	US-09-615-192A-245	Sequence 245, App
42	12.4	82.7	1670	US-09-092-315-4	Sequence 4, Appli
43	12.4	82.7	1764	US-08-918-723-2	Sequence 2, Appli
44	12.4	82.7	1764	US-09-237-507-2	Sequence 2, Appli
45	12.4	82.7	1946	US-09-029-755C-4	Sequence 4, Appli

ALIGNMENTS

```

RESULT 1
US-08-918-148-1
: Sequence 1, Application US/08918148A
: Patent No. 6342220
: GENERAL INFORMATION:
: APPLICANT: Adams, Camellia
: APPLICANT: W.
: APPLICANT: Carter, Paul J.
: APPLICANT: Fendly, Brian M.
: APPLICANT: Gurney, Austin L.
: TITLE OF INVENTION: Agonist Antibodies
: FILE REFERENCE: P0979
: CURRENT APPLICATION NUMBER: US/08/918,148A
: CURRENT FILING DATE: 1997-08-25
: NUMBER OF SEQ ID NOS: 79
: SEQ ID NO 1
: LENGTH: 15
: TYPE: DNA
: ORGANISM: artificial
: FEATURE:
: NAME/KEY: 10F6scfv VH CDR1
: LOCATION: 1-15
: OTHER INFORMATION:
US-08-918-148-1

Query Match      100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCTCTTGATCGGC 15
Db      1 ACCTCTTGATCGGC 15

RESULT 2
US-08-959-381A-3/c
: Sequence 3, Application US/08959381A
: Patent No. 6048711
: GENERAL INFORMATION:
: APPLICANT: HINDMA, SHUJI
: APPLICANT: FUKUSUMI, SHOJI
: APPLICANT: KANAMATA, YUJI
: TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ratner & Prestia
: STREET: P.O. Box 980
: CITY: Valley Forge
: STATE: PA

```

```

: COUNTRY: USA
: ZIP: 19482
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/959,381A
: FILING DATE: 28-OCT-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 286823/1996
: FILING DATE: 29-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Prestia, Paul F
: REGISTRATION NUMBER: 23,031
: REFERENCE/DOCKET NUMBER: TAK-50003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-407-0700
: TELEFAX: 610-407-0700
: TELEX: 846169
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1443 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-959-381A-3

Query Match          93.3%; Score 14; DB 3; Length 1443;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 CCTCTTGATCGGC 15
Db      138 CCTCTTGATCGGC 125

RESULT 3
US-08-959-381A-4/c
: Sequence 4, Application US/08959381A
: Patent No. 6048711
: GENERAL INFORMATION:
: APPLICANT: HINUMA, SHUJI
: APPLICANT: FUKUSUMI, SHOJI
: APPLICANT: KAWAMATA, YUJI
: TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
: TITLE OF INVENTION: POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ratner & Prestia
: STREET: P.O. Box 980
: CITY: Valley Forge
: STATE: PA
: COUNTRY: USA
: ZIP: 19482
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/959,381A
: FILING DATE: 28-OCT-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 286823/1996
: FILING DATE: 29-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Prestia, Paul F
: REGISTRATION NUMBER: 23,031
```

```

: REFERENCE/DOCKET NUMBER: TAK-50003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-407-0700
: TELEFAX: 610-407-0700
: TELEX: 846169
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1626 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-959-381A-4

Query Match          93.3%; Score 14; DB 3; Length 1626;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 CCTCTTGATCGGC 15
Db      321 CCTCTTGATCGGC 308

RESULT 4
US-08-448-619-2/c
: Sequence 2, Application US/08448619
: Patent No. 6140059
: GENERAL INFORMATION:
: APPLICANT: Schwallier, Manfred
: TITLE OF INVENTION: METHOD FOR THE OBTENTION OF NATIVE
: TITLE OF INVENTION: DOMAINS OF VIRAL MEMBRANE PROTEINS, THEIR USE, ESPECIALLY
: TITLE OF INVENTION: AS VACCINE AGAINST HIV AND THESE NATIVE DOMAINS OF VIRAL
: TITLE OF INVENTION: MEMBRANE PROTEINS THEMSELVES
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hardaway Law Firm
: STREET: P.O. Box 10107 Federal Station
: CITY: Greenville
: STATE: SC
: COUNTRY: USA
: ZIP: 29603-0107
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/448,619
: FILING DATE: 29-SEP-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/DE94/00022
: FILING DATE: 12-JAN-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE P 43 01 017.2
: FILING DATE: 16-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Hardaway III, John B.
: REGISTRATION NUMBER: 26,554
: REFERENCE/DOCKET NUMBER: RPE-01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 864-233-6700
: TELEFAX: 864-233-2284
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 39 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Oligonucleotide"
: US-08-448-619-2
```

Query Match 89.3%; Score 13.4; DB 3; Length 39;
Best Local Similarity 93.3%; Pred. No. 64;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
DB 26 ACCTCTTGACCGGC 12

RESULT 5
US-08-448-619-3
Sequence 3, Application US/08448619
Patent No. 6140059
GENERAL INFORMATION:
APPLICANT: Schwallert, Manfred
TITLE OF INVENTION: METHOD FOR THE OBTENTION OF NATIVE
TITLE OF INVENTION: DOMAINS OF VIRAL MEMBRANE PROTEINS, THEIR USE, ESPECIALLY
TITLE OF INVENTION: AS VACCINE AGAINST HIV, AND THESE NATIVE DOMAINS OF VIRAL
TITLE OF INVENTION: MEMBRANE PROTEINS THEMSELVES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hardaway Law Firm
STREET: P. O. Box 10107 Federal Station
CITY: Greenville
STATE: SC
COUNTRY: USA
ZIP: 29603-0107
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,619
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DE94/00022
FILING DATE: 12-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 43 01 017.2
FILING DATE: 16-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hardaway III, John B.
REGISTRATION NUMBER: 26,554
REFERENCE/DOCKET NUMBER: RPE-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 864-233-6700
TELEFAX: 864-233-2284
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Oligonucleotide"
US-08-448-619-3

Query Match 89.3%; Score 13.4; DB 3; Length 39;
Best Local Similarity 93.3%; Pred. No. 64;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
DB 18 ACCTCTTGACCGGC 32

APPLICANT: HINTZ, WILLIAM E.
APPLICANT: LAGOSKY, PETER A.
TITLE OF INVENTION: PRODUCTION OF HETEROLOGOUS PROTEINS IN
TITLE OF INVENTION: FILAMENTOUS FUNGI
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE
STREET: 8TH FLOOR, 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,958
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MITCHARD, LEONARD C.
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 1459-6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NXN UR
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-470-958-11

Query Match 89.3%; Score 13.4; DB 1; Length 60;
Best Local Similarity 93.3%; Pred. No. 66;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
DB 46 ACCTCTTGACCGGC 32

RESULT 7
US-08-321-474-10/C
Sequence 10, Application US/08321474
Patent No. 5710021
GENERAL INFORMATION:
APPLICANT: Hintz, William E.
TITLE OF INVENTION: PRODUCTION OF HETEROLOGOUS PROTEINS IN
TITLE OF INVENTION: FILAMENTOUS FUNGI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye
STREET: 8th Floor, 1100 No. 5710021th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,474
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/988,778
FILING DATE: 10-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C.
REGISTRATION NUMBER: 29009
REFERENCE/DOCKET NUMBER: 617-13
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-321-474-10

Query Match 89.3%; Score 13.4; DB 1; Length 360;
Best Local Similarity 93.3%; Pred. No. 76;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
DB 106 ACCTCTTGATCGGC 92

RESULT 8
US-08-841-349-11/c
Sequence 11, Application US/08841349B
Patent No. 5955564
GENERAL INFORMATION:
APPLICANT: MISHRA, LOPA
TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
FILE REFERENCE: XX/PO4470050
CURRENT APPLICATION NUMBER: US/08/841,349B
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 1177
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
US-08-841-349-11

Query Match 89.3%; Score 13.4; DB 2; Length 1177;
Best Local Similarity 93.3%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
DB 836 ACCTCTTGATCGGC 822

RESULT 9
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: TUBERCULOSIS
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 89.3%; Score 13.4; DB 4; Length 4403765;
Best Local Similarity 93.3%; Pred. No. 63;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
DB 2279627 ACCTCTTGATCGGC 2279641

RESULT 10
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: TUBERCULOSIS
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 89.3%; Score 13.4; DB 4; Length 4411529;
Best Local Similarity 93.3%; Pred. No. 63;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
DB 2277328 ACCTCTTGATCGGC 2277342

RESULT 11
US-08-880-179-1
Sequence 1, Application US/08880179
Patent No. 6091004
GENERAL INFORMATION:
APPLICANT: Ryals, John
APPLICANT: Delaney, Terry
APPLICANT: Friedrich, Leslie
APPLICANT: Weymann, Kristianna
APPLICANT: Lawton, Kay
APPLICANT: Ellis, Daniel
APPLICANT: Unnes, Scott
APPLICANT: Jesse, Taco
APPLICANT: Vos, Pieter
TITLE OF INVENTION: GENE ENCODING A PROTEIN INVOLVED IN THE
TITLE OF INVENTION: SIGNAL TRANSDUCTION CASCADE LEADING TO SYSTEMIC ACQUIRED RE
TITLE OF INVENTION: IN PLANTS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6091004artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880,179
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1909
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9919 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-880-179-1

Query Match 86.7%; Score 13; DB 3; Length 9919;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCG 13
Db 8951 ACCTCTTGATCG 8963

RESULT 12
US-08-972-927-2/c
Sequence 2, Application US/08972927
Patent No. 6166290
GENERAL INFORMATION:
APPLICANT: Rea, Philip A
APPLICANT: Lu, Yu-Ping
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
STREET: Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: US
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,927
FILING DATE: 18-NOV-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,040
FILING DATE: 18-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/061,328

FILING DATE: 08-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-1202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 9936 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-972-927-2

Query Match 86.7%; Score 13; DB 4; Length 9936;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCG 13
Db 6598 ACCTCTTGATCG 6586

RESULT 13
US-08-247-907A-3/c
Sequence 3, Application US/08247907A
Patent No. 5639638
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
APPLICANT: CELESTE, Anthony J.
TITLE OF INVENTION: BMP-11 COMPOSITIONS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247,907A
FILING DATE: May 20, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: G15205-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876-1170
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
STRAIN: Human Activin WC
FEATURE:
NAME/KEY: CDS
LOCATION: 28..183
FEATURE:

NAME/KEY: misc.feature
LOCATION: 184..185
OTHER INFORMATION: /note= "two-thirds of codon at end"
OTHER INFORMATION: of partial clone"
US-08-247-907A-3

Query Match 82.7%; Score 12.4; DB 1; Length 213;
Best Local Similarity 92.9%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 CCTCTTGATCGGC 15
Db 83 CCTCTTGATCGGC 70

RESULT 14
US-08-452-772-3/C
Sequence 3, Application US/08452772
Patent No. 5700911
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
APPLICANT: CELESTE, Anthony J.
TITLE OF INVENTION: BMP-11 COMPOSITIONS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,772
FILING DATE: 30-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/247,907
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: G15205-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876-1170
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
STRAIN: Human Activin WC
FEATURE:
NAME/KEY: CDS
LOCATION: 28..183
FEATURE:
NAME/KEY: misc.feature
LOCATION: 184..185
OTHER INFORMATION: /note= "two-thirds of codon at end"
OTHER INFORMATION: of partial clone"
US-08-452-772-3

Query Match 82.7%; Score 12.4; DB 1; Length 213;
Best Local Similarity 92.9%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
Db 83 CCTCTTGATCGGC 70

RESULT 15
US-09-414-234-3/C
Sequence 3, Application US/09414234
Patent No. 6340668
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
APPLICANT: CELESTE, Anthony J.
THIES, R. Scott
TITLE OF INVENTION: BMP-11 COMPOSITIONS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/414,234
FILING DATE: 07-Oct-1999
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MEINERT, M.C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: G15205-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876-1170
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
STRAIN: Human Activin WC
FEATURE:
NAME/KEY: CDS
LOCATION: 28..183
FEATURE:
NAME/KEY: misc.feature
LOCATION: 184..185
OTHER INFORMATION: /note= "two-thirds of codon at end"
OTHER INFORMATION: of partial clone"
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-414-234-3

Query Match 82.7%; Score 12.4; DB 4; Length 213;
Best Local Similarity 92.9%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 CCTCTTGATCGGC 15
Db 83 CCTCTTGATCGGC 70

Search completed: June 26, 2003, 04:52:10
Job time : 41.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-1

Perfect score: 15
Sequence: 1 accctcttgatcgcc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	14	93.3	1446	9 US-10-083-168-17	Sequence 17, Appl
C 2	14	93.3	1446	9 US-10-083-168-86	Sequence 86, Appl
C 3	14	93.3	1446	9 US-10-251-385-81	Sequence 81, Appl
C 4	14	93.3	1446	9 US-10-251-385-207	Sequence 207, Appl
C 5	14	93.3	1728	10 US-09-826-508-1	Sequence 1, Appl
C 6	13.4	89.3	375	9 US-09-738-626-1488	Sequence 1488, Ap
C 7	13.4	89.3	390	10 US-09-867-701-1577	Sequence 1577, Ap
C 8	13.4	89.3	466	10 US-09-205-658-316	Sequence 316, App
C 9	13.4	89.3	593	10 US-09-998-598-1600	Sequence 1600, Ap
C 10	13.4	89.3	1350	9 US-09-738-626-252	Sequence 252, App
C 11	13.4	89.3	3239	12 US-10-044-090-438	Sequence 438, App
C 12	13.4	89.3	7263	9 US-10-037-182-19	Sequence 19, Appl
C 13	13.4	89.3	7554	9 US-10-037-182-17	Sequence 17, Appl
C 14	13.4	89.3	16489	9 US-09-764-868-1483	Sequence 1483, Ap
C 15	13.4	89.3	3309400	9 US-09-738-626-1	Sequence 1, Appl
C 16	13	86.7	2144	10 US-09-821-839-1	Sequence 1, Appl
C 17	13	86.7	3970	10 US-09-821-839-4	Sequence 4, Appl
C 18	13	86.7	9919	12 US-10-079-035-1	Sequence 1, Appl
C 19	12.4	82.7	25	9 US-10-098-263B-58767	Sequence 58767, A

C 20	12.4	82.7	25	9 US-10-098-263B-95238	Sequence 95238, A
C 21	12.4	82.7	186	10 US-09-783-590-2255	Sequence 2255, Ap
C 22	12.4	82.7	213	9 US-10-029-016-3	Sequence 3, Appl
C 23	12.4	82.7	300	10 US-09-294-093B-4484	Sequence 4484, Ap
C 24	12.4	82.7	306	10 US-09-294-093B-5109	Sequence 5109, Ap
C 25	12.4	82.7	307	9 US-10-015-219-175	Sequence 175, App
C 26	12.4	82.7	307	10 US-09-777-564-175	Sequence 175, App
C 27	12.4	82.7	316	10 US-09-864-761-29857	Sequence 29857, A
C 28	12.4	82.7	318	7 US-08-781-986A-4348	Sequence 4348, Ap
C 29	12.4	82.7	327	9 US-10-071-499A-10	Sequence 10, Appl
C 30	12.4	82.7	334	9 US-09-796-697-9041	Sequence 9041, Ap
C 31	12.4	82.7	334	9 US-10-040-862-9041	Sequence 9041, Ap
C 32	12.4	82.7	333	9 US-10-083-357-16	Sequence 16, Appl
C 33	12.4	82.7	336	9 US-09-970-966-167	Sequence 167, App
C 34	12.4	82.7	396	10 US-09-825-294-167	Sequence 8, Appl
C 35	12.4	82.7	401	10 US-09-998-598-8	Sequence 1978, Ap
C 36	12.4	82.7	405	9 US-09-764-891-1978	Sequence 927, App
C 37	12.4	82.7	409	10 US-09-960-352-927	Sequence 34920, A
C 38	12.4	82.7	411	9 US-09-918-995-34920	Sequence 7377, Ap
C 39	12.4	82.7	417	9 US-09-918-995-7377	Sequence 327, App
C 40	12.4	82.7	422	10 US-09-728-445-327	Sequence 467, App
C 41	12.4	82.7	445	9 US-09-871-161-467	Sequence 23638, A
C 42	12.4	82.7	475	9 US-09-918-995-23638	Sequence 26811, A
C 43	12.4	82.7	487	9 US-09-918-995-26811	Sequence 14852, A
C 44	12.4	82.7	495	9 US-09-918-995-14852	Sequence 588, App
C 45	12.4	82.7	498	10 US-09-954-456-588	

ALIGNMENTS

RESULT 1
US-10-083-168-17/c
Sequence 17, Application US/10083168
Publication No. US20030023069A1
GENERAL INFORMATION:
APPLICANT: Liaw, Chen W.
APPLICANT: Chalmers, Derek T.
APPLICANT: Behan, Dominic P.
APPLICANT: Maciejewski-Lenior, Dominique
APPLICANT: Leonard, James N.
APPLICANT: Ortuno, Daniel
APPLICANT: Lin, I-Lin
TITLE OF INVENTION: Endogenous And No. US20030023069A1-Endogenous, Constitutively
FILE REFERENCE: AREN-0320
CURRENT APPLICATION NUMBER: US/10/083,168
CURRENT FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PatentIn version 3.1
SEQ ID NO 17
LENGTH: 1446
TYPE: DNA
ORGANISM: Homo sapiens
US-10-083-168-17

Query Match 93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
DB 138 CCTCTTGATCGGC 125

RESULT 2
US-10-083-168-86/c
Sequence 86, Application US/10083168
Publication No. US20030023069A1
GENERAL INFORMATION:
APPLICANT: Liaw, Chen W.
APPLICANT: Chalmers, Derek T.
APPLICANT: Behan, Dominic P.

```

; APPLICANT: Maciejewski-Leonior, Dominique
; APPLICANT: Leonard, James N.
; APPLICANT: Ortuno, Daniel
; APPLICANT: Lin, I-Lin
; TITLE OF INVENTION: Endogenous And No. US20030023069A1-Endogenous, Constitutively Act
; FILE REFERENCE: AREN-0320
; CURRENT APPLICATION NUMBER: US/10/083,168
; CURRENT FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 86
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: No. US20030023069A1el Sequence
US-10-083-168-86

Query Match          93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 CCTCTTGATCGGC 15
        |||||||
Db      138 CCTCTTGATCGGC 125

RESULT 3
US-10-251-385-81/c
; Sequence 81, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-81

Query Match          93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 CCTCTTGATCGGC 15
        |||||||
Db      138 CCTCTTGATCGGC 125

RESULT 4
US-10-251-385-207/c
; Sequence 207, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; FILE REFERENCE: AREN-0040
; PRIOR APPLICATION NUMBER: US/10-040
; PRIOR FILING DATE: 2000-04-07
```

```

; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 207
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-207

Query Match          93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 CCTCTTGATCGGC 15
        |||||||
Db      138 CCTCTTGATCGGC 125

RESULT 5
US-09-826-508-1/c
; Sequence 1, Application US/09826508
; Patent No. US20010025099A1
; GENERAL INFORMATION:
; APPLICANT: Nabil Elshourbagy
; APPLICANT: Lisa Yawler
; TITLE OF INVENTION: G-Protein-Coupled Receptor Polypeptides
; TITLE OF INVENTION: and Polynucleotides
; FILE REFERENCE: GP-70744USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-826-508-1

Query Match          93.3%; Score 14; DB 10; Length 1728;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 CCTCTTGATCGGC 15
        |||||||
Db      287 CCTCTTGATCGGC 274

RESULT 5
US-09-738-626-1488
; Sequence 1488, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: YATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
```

PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-06-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1488
LENGTH: 375
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1488

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
Db 60 ACCTCTTGATCGGC 74

RESULT 7
US-09-867-701-1577
Sequence 1577, Application US/09867701
Patent No. US2002013237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867, 701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1577
LENGTH: 390
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(390)
OTHER INFORMATION: n = A,T,C or G
US-09-867-701-1577

Query Match 89.3%; Score 13.4; DB 10; Length 390;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
Db 84 ACCTCTTGATCGGC 98

RESULT 8
US-09-205-658-316/C
Sequence 316, Application US/09205658
Patent No. US20010029617A1
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Ogg, Scott
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
FILE REFERENCE: 00786/351004
CURRENT APPLICATION NUMBER: US/09/205,658
CURRENT FILING DATE: 1998-12-03
EARLIER APPLICATION NUMBER: 08/857,076
EARLIER FILING DATE: 1997-05-15
EARLIER APPLICATION NUMBER: 08/888,534
EARLIER FILING DATE: 1997-07-07
EARLIER APPLICATION NUMBER: US98/10080
EARLIER FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 328
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 316
LENGTH: 466
TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-09-205-658-316

Query Match 89.3%; Score 13.4; DB 10; Length 466;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
Db 258 ACCTCTTGATCGGC 244

RESULT 9
US-09-998-598-1600/C
Sequence 1600, Application US/09998598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 1600
LENGTH: 593
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-1600

Query Match 89.3%; Score 13.4; DB 10; Length 593;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
Db 373 ACCTCTTGATCGGC 359

RESULT 10
US-09-738-626-252/C
Sequence 252, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 252
LENGTH: 1350
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-252

Query Match 89.3%; Score 13.4; DB 9; length 1350;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
|||||
Db 113 ACCTCTGGATCGGC 99

RESULT 11
US-10-044-090-438
Sequence 438, Application US/10044090
Patent No. US20020137081A1
GENERAL INFORMATION:
APPLICANT: Olga Bandman
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
FILE REFERENCE: PA-0028 US
CURRENT APPLICATION NUMBER: US/10/044,090
CURRENT FILING DATE: 2002-01-09
NUMBER OF SEQ ID NOS: 850
SOFTWARE: PERL Program
SEQ ID NO 438
LENGTH: 3239
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020137081A1 241624.1
US-10-044-090-438

Query Match 89.3%; Score 13.4; DB 12; length 3239;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
|||||
Db 1958 ACCTCTGGATCGGC 1972

RESULT 12
US-10-037-182-19
Sequence 19, Application US/10037182
Publication No. US20030044899A1
GENERAL INFORMATION:
APPLICANT: Trygvsason, Karl
APPLICANT: Doi, Masayuki
TITLE OF INVENTION: Recombinant Laminin 10
FILE REFERENCE: 99-274-F
CURRENT APPLICATION NUMBER: US/10/037,182
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,449
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/279,282
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 19
LENGTH: 7263
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(4716)
US-10-037-182-19

Query Match 89.3%; Score 13.4; DB 9; length 7263;

Best Local Similarity 93.3%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
|||||
Db 4445 ACCTCTGGATCGGC 4459

RESULT 13
US-10-037-182-17
Sequence 17, Application US/10037182
Publication No. US20030044899A1
GENERAL INFORMATION:
APPLICANT: Trygvsason, Karl
APPLICANT: Doi, Masayuki
TITLE OF INVENTION: Recombinant Laminin 10
FILE REFERENCE: 99-274-F
CURRENT APPLICATION NUMBER: US/10/037,182
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,449
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/279,282
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 17
LENGTH: 7554
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (193)..(5007)
NAME/KEY: sig-peptide
LOCATION: (193)..(291)
US-10-037-182-17

Query Match 89.3%; Score 13.4; DB 9; length 7554;
Best Local Similarity 93.3%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
|||||
Db 4736 ACCTCTGGATCGGC 4750

RESULT 14
US-09-764-868-1483
Sequence 1483, Application US/09764868
Patent No. US20020168711A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764,868
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1483
LENGTH: 16489
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-868-1483

Query Match 89.3%; Score 13.4; DB 9; length 16489;
Best Local Similarity 93.3%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
|||||
Db 16177 ACCTCTGGATCGGC 16191

RESULT 15

US-09-738-626-1/c

; Sequence 1, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAMA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAOKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 1

; LENGTH: 3309400

; TYPE: DNA

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-1

Query Match 89.3%; Score 13.4; DB 9; Length 3309400;

Best Local Similarity 93.3%; Pred. No. 1.9e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15

|||||

Db 238288 ACCTCTTGATCGGC 238274

Search completed: June 26, 2003, 04:56:43

Job time : 75.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-1
Perfect score: 15
Sequence: 1 accctcttgatcgcc 15

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	237	14	BQ757051 EBem10_SQ
2	15	100.0	433	13	BK376561 EBem05_SQ
3	15	100.0	441	9	AJ434075 AJ434075
4	15	100.0	444	14	BQ461961 HD02115r
5	15	100.0	458	13	BT777123 EB003_SQ
6	15	100.0	474	14	BQ468011 HRO1K11r

C 7	15	100.0	518	12	BG300356	BG300356 HVSMEB001
C 8	15	100.0	521	13	BT780309	BT780309 EBem06_SQ
C 9	15	100.0	558	14	AV941160	AV941160 AV941160
C 10	15	100.0	565	14	BQ465771	BQ465771 HU04K09r
C 11	15	100.0	575	14	BQ471255	BQ471255 HV01N14r
C 12	15	100.0	597	14	BQ464764	BQ464764 HU01G04T
C 13	15	100.0	600	17	AV915369	AV915369 AV915369
C 14	15	100.0	606	17	BH474182	BH474182 BOCVM88TF
C 15	15	100.0	607	17	BQ465677	BQ465677 HU04F22r
C 16	15	100.0	615	14	BQ471432	BQ471432 HV02G15r
C 17	15	100.0	620	13	BM099080	BM099080 EB0105_SQ
C 18	15	100.0	636	12	BG415523	BG415523 HVSMEK000
C 19	15	100.0	653	12	BG415434	BG415434 HVSMEK000
C 20	15	100.0	670	17	BH446702	BH446702 BOCVJ04TR
C 21	15	100.0	671	13	BT954106	BT954106 HVSMEB001
C 22	15	100.0	799	15	BF254300	BF254300 HVSMEF000
C 23	15	100.0	851	12	BG418413	BG418413 HVSMEK002
C 24	15	100.0	932	12	BF623261	BF623261 HVSMEB001
C 25	15	100.0	1037	13	BT950768	BT950768 HVSME1002
C 26	15	100.0	273	10	BE415681	BE415681 MML038.B0
C 27	14	93.3	281	9	AL829578	AL829578 AL829578
C 28	14	93.3	410	12	BF146084	BF146084 WHE1837.A
C 29	14	93.3	475	13	BM027631	BM027631 CSAC30019
C 30	14	93.3	499	13	BT805258	BT805258 S034EB11.S
C 31	14	93.3	512	9	AT773000	AT773000 EST254100
C 32	14	93.3	524	14	BQ273893	BQ273893 rc72c10.y
C 33	14	93.3	537	13	BT669564	BT669564 603294282
C 34	14	93.3	553	14	BO853560	BO853560 OGB20N12.
C 35	14	93.3	568	13	BT302594	BT302594 BT302594
C 36	14	93.3	592	13	BT260486	BT260486 BT260486
C 37	14	93.3	607	12	BT910143	BT910143 602805504
C 38	14	93.3	620	17	A2432572	A2432572 1M0218K01
C 39	14	93.3	621	12	BF530127	BF530127 602040506
C 40	14	93.3	624	13	BT033194	BT033194 BT033194
C 41	14	93.3	632	13	BT604112	BT604112 603240448
C 42	14	93.3	638	13	BT037202	BT037202 BT037202
C 43	14	93.3	639	13	BT034888	BT034888 BT034888
C 44	14	93.3	640	10	BE427382	BE427382 PSR6374-B
C 45	14	93.3	668	13	BT601027	BT601027 603249503

ALIGNMENTS

RESULT 1
BQ757051/c 237 bp mRNA linear EST 26-JUL-2002
LOCUS
DEFINITION
EBem10_SQ002_K07_R embryo, 2 day germination, no treatment, cv
Optic, EBem10 Hordeum vulgare cDNA clone EBem10_SQ002_K07 5', mRNA
sequence.

ACCESSION
BQ757051
VERSION
BQ757051.1 GI:21965523
KEYWORDS
EST.
SOURCE
Hordeum vulgare.
ORGANISM
Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae
; Triticeae; Hordeum.

REFERENCE
1 (bases 1 to 237)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

TITLE
Development of Barley Transcriptome Resources
Unpublished (2001)
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk.

COMMENT

JOURNAL

FEATURES

source
1..237
/organism="Hordeum vulgare"

BASE COUNT	106 a	103 c	115 g	109 t
ORIGIN	45 a	68 c	73 g	51 t
Query Match	100.0%;	Score 15;	DB 14;	Length 237;
Best Local Similarity	100.0%;	Prod. No. 6,1e+02;		
Matches	15;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;
Oy	1	ACCTCTGGATCGC	15	
Db	126	ACCTCTGGATCGC	112	
RESULT 2				
BM376561/c				
LOCUS	BM376561	433 bp	mRNA	linear
DEFINITION	EBem05_S0002_J05_R embryo, 14 DPA, no treatment, cv Optic, EBem05			
ACCESSION	BM376561			
VERSION	BM376561.2			
KEYWORDS	GI:21934181			
SOURCE	EST.			
ORGANISM	Hordeum vulgare.			
	Hordeum vulgare.			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae			
	; Triticeae; Hordeum.			
REFERENCE	1 (bases 1 to 433)			
AUTHORS	Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,			
TITLE	Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.			
JOURNAL	Development of Barley Transcriptome Resources			
COMMENT	Unpublished (2001)			
	On Jan 10, 2002 this sequence version replaced gi:18119951.			
	Contact: Waugh R, Marshall DF			
	Genome Dynamics/Computational Biology			
	Scottish Crop Research Institute			
	Invergowrie, Dundee, DD2 5DA, Scotland, UK			
	Tel: 00 44 1382 562731			
	Fax: 00 44 1382 562426			
	Email: est@scri.sari.ac.uk			
	All sequence has a Phred quality score of 20 or over			
	Seq primer: M13 reverse.			
FEATURES	Location/Qualifiers			
source	1..433			
	/organism="Hordeum vulgare"			
	/cultivar="Optic"			
	/db_xref="taxon:4513"			
	/clone="EBem05_S0002_J05"			
	/clone_1="EBem05_S0002_J05"			
	/clone_1ib="embryo, 14 DPA, no treatment, cv Optic,			
	EBem05"			
	/tissue_type="embryo"			
	/dev_stage="14 DPA"			
	/lab_host="DH10B"			
	/note="Vector: pSPOR1; Site_1: Sal I; Site_2: Not I;			
	Non-normalised library, directionally cloned into pSPOR1			
	Derived from embryos dissected from germinating grains (2			
	day) in glasshouse grown barley plants. Developed as part			
	of the barley transcriptome resources of BBSRC/SEERAD			
	funded cereal IGF (Investigating Gene Function) project."			
	BBSRC/SEERAD funded cereal IGF (Investigating Gene			
	Function) project."			

[illegible]

Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 444 Std Error: 0.00
 Plate: 2 row: 1 column: 15
 Seq primer: M13rev.

FEATURES

SOURCE

1. .444
 /organism="Hordeum vulgare"
 /cultivar="Golden Promise"
 /db_xref="taxon:4513"
 /clone_1db="HD02L15"
 /clone_1lb="HD"
 /tissue-type="callus"
 /dev_stage="callus (5-10 mm in diameter)"
 /lab_host="XL10-Gold"
 /note="Vector: Bluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter sites upstream of the cloning. To excise the insert, restriction sites upstream of the cloning should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1 kb"

BASE COUNT 105 a 114 c 121 g 104 t

ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 444;
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
 |||||||
 Db 153 ACCTCTTGATCGGC 139

RESULT 5

BI777123/c 458 bp mRNA linear EST 23-JUL-2002
 LOCUS EBR003_S0002_A18_R root, 3 week, waterlogged, cv Optic, EBR003
 DEFINITION Hordeum vulgare cDNA clone EBR003_S0002_A18 5', mRNA sequence.
 BI777123
 BI777123.2 GI:21948259
 EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 458)
 Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D. F. M. and Waugh, R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)
 On Sep 26, 2001 this sequence version replaced gi:15780015.
 Contact: Waugh R, Marshall DF
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.

FEATURES

SOURCE

1. .458
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="EBR003_S0002_A18"
 /clone_1db="root, 3 week, waterlogged, cv Optic, EBR003"

/tissue-type="root"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /note="Vector: pSPORT1, Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from roots of 3 week old waterlogged barley
 plants. Developed as part of the barley transcriptome
 resources of BBSRC/SERAD funded cereal IGF (Investigating
 Gene Function) project."

BASE COUNT 111 a 116 c 121 g 110 t

ORIGIN

Query Match 100.0%; Score 15; DB 13; Length 458;
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
 |||||||
 Db 153 ACCTCTTGATCGGC 139

RESULT 6

BQ468011/c 474 bp mRNA linear EST 30-MAY-2002
 LOCUS HR01K11r HR Hordeum vulgare cDNA clone HR01K11 5-PRIME, mRNA
 DEFINITION sequence.
 BQ468011
 BQ468011.1 GI:21275793
 EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 474)
 Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
 EST sequencing and analysis in barley (2002)
 Unpublished (2002)
 Contact: Stein N11s
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 474 Std Error: 0.00
 Plate: 1 row: 1 column: 11
 Seq primer: M13rev.

FEATURES

SOURCE

1. .474
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HR01K11"
 /clone_1db="HR"
 /tissue-type="roots"
 /dev_stage="grown on filter paper at 25 Grad C for 48
 hours in petridish, 1-1.5 cm in length"
 /lab_host="XL10-Gold"
 /note="Vector: Bluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter sites upstream of the cloning. To excise the insert, restriction sites upstream of the cloning should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 800 bp"

BASE COUNT 114 a 120 c 130 g 110 t

ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 474;
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;

Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 QY 1 ACCTCTGGATCGGC 15
 |||||||||||||
 Db 153 ACCTCTGGATCGGC 139

RESULT 7
 BG300356/c 518 bp mRNA linear EST 22-OCT-2001
 LOCUS HVSMEB0012021f Hordeum vulgare seedling shoot EST library
 DEFINITION HVCNMA0002 (Dehydration stress) Hordeum vulgare cDNA clone
 HVSMEB0012021f, mRNA sequence.
 BG300356
 BG300356
 BG300356.1 GI:13097883
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Hordeum vulgare.
 Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 ; Triticeae; Hordeum.
 1 (bases 1 to 518)
 Wing, R., Close, T. J., Kleinbofs, A., Wise, R., Begum, D., Frisch, D., Yu
 , Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D. W., Fenton
 , R. D., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex drought-stressed seedling shoot cDNA
 library
 Unpublished (2001)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 235
 Seq primer: AATTAACCTCCTCAATAGGCG
 High quality sequence stop: 411.
 Location/Qualifiers
 1. 518
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEB0012021f"
 /clone.lib="Hordeum vulgare seedling shoot EST library
 HVCNMA0002 (Dehydration stress)"
 /tissue.type="Seedling shoot"
 /lab_host="TJCI21"
 /note="vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper
 with water, nystatin and cefotaxime in covered
 crystallization dishes. Five-day old seedlings were
 incubated at 90% RH for 24 hr. Shoots were then harvested,
 total RNA was prepared, poly(A) RNA was purified, one
 primary unamplified cDNA library was made, 600000 pfu were
 in vivo excised to give phagescript SK(-) cDNA phagemids.
 These steps were performed in the TJ Close Laboratory at
 the University of California, Riverside (Choi, Close,
 Fenton). Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders
 see Close TJ, Wing R, Kleinbofs A, Wise R (2001)
 Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/gnpages/bgn/31/cover.html)"

BASE COUNT 135 a 145 c 120 g 117 t 1 others
 ORIGIN
 Query Match 100.0%; Score 15; DB 12; Length 518;
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 ACCTCTGGATCGGC 15
 |||||||||||||
 Db 207 ACCTCTGGATCGGC 193

RESULT 8
 B1780309/c 521 bp mRNA linear EST 23-JUL-2002
 LOCUS Ebem06_S0001_E02.R embryo, 21 DPA, no treatment, cv Optic, Ebem06
 DEFINITION Hordeum vulgare cDNA clone Ebem06_S0001_E02 5', mRNA sequence.
 B1780309
 B1780309.2 GI:21934849
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Hordeum vulgare.
 Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 ; Triticeae; Hordeum.
 1 (bases 1 to 521)
 Hedley, P., Yu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D. F. M. and Maugh, R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)
 On Sep 26, 2001 this sequence version replaced gi:15783201.
 Contact: Maugh R, Marshall DP
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.
 Location/Qualifiers
 1. 521
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="Ebem06_S0001_E02"
 /clone.lib="embryo, 21 DPA, no treatment, cv Optic,
 Ebem06"
 /tissue.type="embryo"
 /lab_host="DH10B"
 /dev_stage="21 DPA"
 /note="vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from embryos dissected from developing grains (21
 days post anthesis) in glasshouse grown barley plants.
 BBSRC/SEERAD funded cereal IGF (Investigating Gene
 Function) project."

BASE COUNT 127 a 121 c 145 g 128 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 13; Length 521;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 ACCTCTGGATCGGC 15
 |||||||||||||
 Db 136 ACCTCTGGATCGGC 122

RESULT 9

AV941160/c 558 bp mRNA linear EST 18-JAN-2002
 LOCUS AV941160 K. Sato unpublished cDNA library, strain H602 adult,
 DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum
 cDNA clone bnh31m09 5', mRNA sequence.

ACCESSION AV941160
 VERSION AV941160.1 GI:18236957
 KEYWORDS Hordeum vulgare subsp. spontaneum.
 SOURCE Hordeum vulgare subsp. spontaneum.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 ; (bases 1 to 558)
 1 (bases 1 to 558)
 Sato,K., Saitoh,D. and Takeda,K.
 Barley EST sequencing project in NIG and Okayama Univ
 Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..558
 /organism="Hordeum vulgare subsp. spontaneum"
 /strain="H602"
 /db_xref="taxon:77009"
 /clone_lib="K. Sato unpublished cDNA library, strain H602
 adult, heading stage top three leaves"
 /tissue_type="top three leaves"
 /dev_stage="adult, heading stage"
 /dev_stage="adult, heading stage"
 142 c 149 g 129 t

BASE COUNT 138 a 142 c 149 g 129 t
 ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 558;
 Best Local Similarity 100.0%; Pred. No. 8e+02; 0; Indels 0; Gaps 0;
 Matches 15; Conservative 0; Mismatches 0;

QY 1 ACCTCTGATCGGC 15
 |||||
 Db 169 ACCTCTGATCGGC 155

RESULT 10 565 bp mRNA linear EST 30-MAY-2002
 BQ465771/c
 LOCUS BQ465771 HU Hordeum vulgare cDNA clone HU04K09 5-PRIME, mRNA
 DEFINITION sequence.
 ACCESSION BQ465771
 VERSION BQ465771.1 GI:21273553
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 ; (bases 1 to 565)
 1 (bases 1 to 565)
 Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
 ,A.
 Barley ESTs from germinating seeds
 Unpublished (2002)
 COMMENT Contact: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 565 Std Error: 0.00
 Plate: 4 row: K column: 9

Seq primer: M13rev.
 Location/Qualifiers
 1..565
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone_lib="HU04K09"
 /clone_lib="HU"
 /tissue_type="germinating seeds"
 /dev_stage="germinating seeds, 16-48 h"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
 cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds,
 16-48h. Due to a cloning artefact caused by the kit, in
 most cases the EcoRI site is NOT present, as well as the
 EcoRI adapter used for cloning. To excise the insert,
 BamHI, SalI, PstI). NOTE: Also due to the cloning system
 used Blue/White selection for recombinants is not 100%
 reliable. Average insert size is 1.2 kb"

BASE COUNT 136 a 134 c 158 g 137 t
 ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 565;
 Best Local Similarity 100.0%; Pred. No. 8.1e+02; 0; Indels 0; Gaps 0;
 Matches 15; Conservative 0; Mismatches 0;

QY 1 ACCTCTGATCGGC 15
 |||||
 Db 148 ACCTCTGATCGGC 134

RESULT 11 575 bp mRNA linear EST 30-MAY-2002
 BQ471255/c
 LOCUS BQ471255 HV Hordeum vulgare cDNA clone HV01N14 5-PRIME, mRNA
 DEFINITION sequence.
 ACCESSION BQ471255
 VERSION BQ471255.1 GI:21279037
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 ; (bases 1 to 575)
 1 (bases 1 to 575)
 Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
 ,A.
 Barley ESTs from germinating seeds
 Unpublished (2002)
 COMMENT Contact: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 575 Std Error: 0.00
 Plate: 1 row: N column: 14
 Seq primer: T3.
 Location/Qualifiers
 1..575
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone_lib="HV01N14"
 /clone_lib="HV"
 /tissue_type="germinating seeds"
 /dev_stage="germinating seeds (48-96 h)"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
 cDNA); Site_2: XhoI (3'-end of cDNA); Roots were grown for
 two days on filter paper at room temperature. Due to a

cloning artefact caused by the kit, in most cases the EcoRI site is NOT present as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI).
NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1 kb"

BASE COUNT 140 a 138 c 159 g 138 t

Query Match 100.0%; Score 15; DB 14; Length 575;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
DB 150 ACCTCTTGATCGGC 136

RESULT 12
BO464764/c 597 bp mRNA linear EST 30-MAY-2002
LOCUS H001G04T HU Hordeum vulgare CDNA clone H001G04 5-PRIME. mRNA
DEFINITION sequence.
BO464764
ACCESSION BO464764.1 GI:21272546
VERSION
KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare.
Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 597)
Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and Graner

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Barley ESTs from germinating seeds
Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert length: 597 Std Error: 0.00
Plate: 1 row: 6 column: 4
Seq primer: T3.

FEATURES

Source location/Qualifiers

1..597
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="H001G04"
/clone_1ib="HU"
/tissue_type="germinating seeds"
/dev_stage="germinating seeds, 16-48 h"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK-; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds, 16-48h. Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"

BASE COUNT 149 a 148 c 160 g 140 t

Query Match 100.0%; Score 15; DB 14; Length 597;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
DB 165 ACCTCTTGATCGGC 151

RESULT 13
AV915369/c 600 bp mRNA linear EST 18-JAN-2002
LOCUS AV915369 K. Sato unpublished cdna library, cv. Haruna Nijo
DEFINITION germination shoots Hordeum vulgare subsp. vulgare CDNA clone
bags14b14 5', mRNA sequence.
AV915369
ACCESSION AV915369.1 GI:18211146
VERSION
KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare subsp. vulgare.
Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 600)
Sato, K., Saitoh, D. and Takeda, K.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genetics.nig.ac.jp.
location/Qualifiers
1..600

FEATURES

Source location/Qualifiers
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna Nijo"
/db_xref="taxon:112509"
/clone="bags14b14"
/clone_1ib="K. Sato unpublished cdna library, cv. Haruna Nijo germination shoots"
/tissue_type="shoots"
/dev_stage="germination"
BASE COUNT 149 a 147 c 160 g 142 t 2 others

Query Match 100.0%; Score 15; DB 10; Length 600;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
|||||
DB 163 ACCTCTTGATCGGC 149

RESULT 14
BH474182/c 606 bp DNA linear GSS 13-DEC-2001
LOCUS BOCVM88TF BOCV Brassica oleracea genomic clone BOCVM88, DNA
DEFINITION sequence.
BH474182
ACCESSION BH474182.1 GI:17682293
VERSION
KEYWORDS
SOURCE
ORGANISM

Brassica oleracea.
Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 606)
Town, C.D., Van Aken, S., Ullterback, T. and Fraser, C.M.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOCVM88TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES

source

1.606
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOGVM88"
 /clone_1lb="BOGV"
 /note="Vector: pHOSt1; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOSt1 using BstXI linkers"

BASE COUNT 150 a 148 c 128 g 180 t
 ORIGIN

Query Match 100.0%; Score 15; DB 17; Length 606;
 Best Local Similarity 100.0%; Pred. No. 8.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
 ||||||||||||
 Db 330 ACCTCTGGATCGGC 316

RESULT 15
 BQ465677/c 607 bp mRNA linear EST 30-MAY-2002
 LOCUS HU04F22r HU Hordeum vulgare cDNA clone HU04F22 5-PRIME, mRNA
 DEFINITION

ACCESSION BQ465677
 VERSION BQ465677.1 GI:21273459
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.

1 (bases 1 to 607)
 Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
 ,A.

Barley ESTs from germinating seeds
 Unpublished (2002)
 TITLE JOURNAL
 COMMENT

Contact: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595

Email: stein@ipk-gatersleben.de
 Insert Length: 607 Std Error: 0.00
 Plate: 4 row: F column: 22
 Seq primer: M13rev.

FEATURES

source

1.607
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HU04F22"
 /clone_1lb="HU"
 /tissue.type="germinating seeds"
 /dev_stage="germinating seeds, 16-48 h"
 /lab_host="Xli0-Gold"
 /note="Vector: pBluescript SK+, Site_1: EcoRI (5'-end of
 cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds,
 16-48h. Due to a cloning artefact caused by the kit, in
 most cases the EcoRI site is NOT present, as well as the
 EcoRI adapter used for cloning. To excise the insert,
 restriction sites upstream EcoRI should be used (e.g.
 BamHI, SalI, PstI). NOTE: Also due to the cloning system
 used Blue/white selection for recombinants is not 100%

BASE COUNT 153 a 154 c 161 g 139 t
 ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 607;
 Best Local Similarity 100.0%; Pred. No. 8.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTGGATCGGC 15
 ||||||||||||
 Db 179 ACCTCTGGATCGGC 165

Search completed: June 26, 2003, 04:50:05
 Job time : 754.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds

(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-13

Perfect score: 15

Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	6	AR183480	AR183480 Sequence
2	15	100.0	165	9	HS1MHCV10	Z29985 H.sapiens D
3	15	100.0	217	9	HSJ227738	AJ227738 Homo sapi
4	15	100.0	223	9	AF017465	AF017465 Homo sapi
5	15	100.0	223	9	HSJ227734	AJ227734 Homo sapi
6	15	100.0	246	9	HSJ306240	AJ306240 Homo sapi
7	15	100.0	255	9	AF297164	AF297164 Homo sapi
8	15	100.0	258	9	HSEM12V3	Y12032 H.sapiens r
9	15	100.0	261	9	HSJ406644	AJ406644 Homo sapi
10	15	100.0	273	9	HSJ415895	AJ415895 Homo sapi
11	15	100.0	278	9	HSJ77321	Z77321 H.sapiens g
12	15	100.0	279	9	AF077461	AF077461 Homo sapi
13	15	100.0	279	9	AT013309	AT013309 Homo sapi
14	15	100.0	283	9	AT003824	AT003824 Homo sapi
15	15	100.0	291	9	HUMIGL2E5	L04325 Human (cell
16	15	100.0	291	9	HSJ80382	Z80382 H.sapiens B
17	15	100.0	293	9	HSJG14	X67073 H.sapiens I
18	15	100.0	294	9	AF119800	AF119800 Homo sapi
19	15	100.0	294	9	AF119801	AF119801 Homo sapi
20	15	100.0	294	9	AF119802	AF119802 Homo sapi
21	15	100.0	294	9	AF119803	AF119803 Homo sapi
22	15	100.0	294	9	AF119804	AF119804 Homo sapi
23	15	100.0	294	9	AF151702	AF151702 Homo sapi
24	15	100.0	294	9	AF151704	AF151704 Homo sapi
25	15	100.0	294	9	HSJ09062	HSJ09062 Human Immun
26	15	100.0	294	9	HSJGVH38	X92267 H.sapiens g
27	15	100.0	294	9	HSJGDP35	Z12337 H.sapiens g
28	15	100.0	296	6	AXJ470326	AXJ470326 Sequence
29	15	100.0	297	9	HUMIGHNK	L03824 Human IG re
30	15	100.0	297	9	HSJ80419	Z80419 H.sapiens B
31	15	100.0	300	9	HSJGDP34	Z12336 H.sapiens g
32	15	100.0	300	9	HUMIGHNL	L03825 Human IG re
33	15	100.0	300	9	HSJ80669	Z80669 H.sapiens B
34	15	100.0	303	9	HSJ80625	Z80625 H.sapiens B
35	15	100.0	304	9	AF077475	AF077475 Homo sapi
36	15	100.0	304	9	AF077481	AF077481 Homo sapi
37	15	100.0	305	9	AF077428	AF077428 Homo sapi
38	15	100.0	306	9	AF077456	AF077456 Homo sapi
39	15	100.0	306	9	HSB83408N	X87074 H.sapiens B
40	15	100.0	307	9	AF077478	AF077478 Homo sapi
41	15	100.0	309	9	AF077466	AF077466 Homo sapi
42	15	100.0	309	9	HSB83607N	X87070 H.sapiens B
43	15	100.0	311	9	HSJ289464	AJ289464 Homo sapi
44	15	100.0	312	9	AF077455	AF077455 Homo sapi
45	15	100.0	312	9	AF077471	AF077471 Homo sapi

ALIGNMENTS

RESULT 1
AR183480
LOCUS AR183480
DEFINITION Sequence 13 from patent US 6342220.
ACCESSION AR183480
VERSION AR183480.1 GI:20227449
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 13 29-JAN-2002;
FEATURES Location/Qualifiers

```

source
1..15
/organism="unknown"
BASE COUNT      5 a      4 c      3 g      3 t
ORIGIN
Query Match      100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. NO. 1.4e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
       |||
Db      1 GACTACTACATGAGC 15

RESULT 2
HSTMHCY10      165 bp      DNA      linear      PRI 21-JUN-1994
LOCUS
DEFINITION
H.sapiens DNA for immunoglobulin heavy chain variable region
(DA-10).
ACCESSION
229985
VERSION
229985.1 GI:505455
KEYWORDS
immunoglobulin heavy chain variable region.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 165)
Cook,G.P., Tomlinson,I.M., Walter,G., Carter,N.P., Rietman,H.,
Bulwela,L., Winter and Rabblits,T.H.
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 165)
Tomlinson,M.
Direct Submission
Submitted (10-FEB-1994) Ian Tomlinson, MRC Centre for Protein
Engineering, Hills Road, Cambridge, CB2 2QH, U.K
3 (bases 1 to 165)
Cook,G.P., Tomlinson,I.M., Walter,G., Rietman,H., Carter,N.P.,
Bulwela,L., Winter,G. and Rabblits,T.H.
A map of the human immunoglobulin VH locus completed by analysis of
the telomeric region of chromosome 14q
JOURNAL
MEDLINE
95004581
PUBMED
7920635
FEATURES
source
Location/Qualifiers
1..165
/organism="Homo sapiens"
/isolate="der 8 chromosome from 'DAUDI' cell line"
/db_xref="taxon:9606"
/dev_stage="adult"
/germline
1..165
/pseudo
/codon_start=1
/product="immunoglobulin heavy chain variable region
(DA-10)"
/db_xref="PDB:e72321"
BASE COUNT      51 a      36 c      44 g      34 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 165;
Best Local Similarity 100.0%; Pred. NO. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
       |||
Db      10 GACTACTACATGAGC 24

RESULT 3
HSJ227738      217 bp      DNA      linear      PRI 10-MAR-1998
LOCUS
DEFINITION
Homo sapiens DNA for rearranged immunoglobulin heavy chain gene,
4-1YH3.

```

```

ACCESSION
AJ227738
VERSION
AJ227738.1 GI:2951653
KEYWORDS
heavy chain; immunoglobulin; variable region.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 217)
Delabie,J.
Direct Submission
Submitted (03-MAR-1998) Delabie J., Pathology, University of
Leuven, Minderbroedersstraat 12, B-3000, BELGIUM
2 (bases 1 to 217)
Delabie,J. and Tiersen,A.
Evidence for clonal expansion and somatic hypermutations of the
marginal zone B cells in the lymph node and spleen
Unpublished
JOURNAL
FEATURES
source
Location/Qualifiers
1..217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/rearranged
<1..>217
/codon_start=2
/product="immunoglobulin heavy chain variable region"
/protein_id="CA12716.1"
/db_xref="GI:2951654"
/translation="FTFSDYYMSWIRQAPGKLEWLSDMSSSGTSIYYADSVKRGPTI
SRDNKNSVYQWMSLRADPTAVYHCAR"
<1..>217
/product="immunoglobulin heavy chain variable region"
BASE COUNT      58 a      55 c      59 g      45 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. NO. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
       |||
Db      14 GACTACTACATGAGC 28

RESULT 4
AF017465      223 bp      mRNA      linear      PRI 18-OCT-1997
LOCUS
DEFINITION
Homo sapiens isolate case 5 rearranged immunoglobulin heavy chain
variable region (Igh) mRNA, partial cds.
ACCESSION
AF017465
VERSION
AF017465.1 GI:2547109
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 223)
Delabie,J. and Pittaluga,S.
Direct Submission
Submitted (08-AUG-1997) Pathology, University of Leuven,
Minderbroedersstraat 12, Leuven B-3000, Belgium
JOURNAL
FEATURES
source
Location/Qualifiers
1..223
/organism="Homo sapiens"
/isolate="case 5"
/db_xref="taxon:9606"
<1..>223
/gene="Igh"
<1..>223
/gene="Igh"
<1..>223
/gene="Igh"
/codon_start=2
/product="rearranged immunoglobulin heavy chain variable
region"
/protein_id="AAB81258.1"

```

```

/db_xref="GI:2547110"
/translation="SGFTFSDYMSWIRQAPGKGLFWYSYSSSSSYTNADSVKGRF
TISRDNKNSLYLQMNSLRAEDTAVYCAR"
BASE COUNT      61 a      54 c      59 g      49 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
        |||
Db      20 GACTACTACATGAGC 34

RESULT 5
HSJ227734      223 bp      DNA      linear      PRI 10-MAR-1998
LOCUS      Homo sapiens DNA for rearranged immunoglobulin heavy chain gene,
DEFINITION      7-9YH3.
ACCESSION      AJ227734.1 GI:2951645
VERSION      heavy chain: immunoglobulin; variable region.
KEYWORDS      Homo sapiens.
SOURCE      Homo sapiens.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 223)
AUTHORS      Delabie,J.
TITLE      Direct Submission
JOURNAL      Submitted (03-MAR-1998) Delabie J., Pathology, University of
              Leuven, Minderbroedersstraat 12, B-3000, BELGIUM
REFERENCE      2 (bases 1 to 223)
AUTHORS      Delabie,J. and Tiersens,A.
TITLE      Evidence for clonal expansion and somatic hypermutations of the
JOURNAL      marginal zone B cells in the lymph node and spleen
FEATURES
source      location/Qualifiers
              1..223
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /rearranged
              <1..>223
              /codon_start=2
              /product="immunoglobulin heavy chain variable region"
              /protein_id="CAI12712.1"
              /db_xref="GI:2951646"
              /translation="SGFTFSDYMSWIRQAPGKGLFWYSYSSSSSYTNADSVKGRF
              TISRDNKNSLYLQMNSLRAEDTAVYCAR"
              <1..>223
              /product="immunoglobulin heavy chain variable region"
V_region
BASE COUNT      61 a      54 c      59 g      49 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
        |||
Db      20 GACTACTACATGAGC 34

RESULT 6
HSA306240      246 bp      DNA      linear      PRI 10-JUN-2001
LOCUS      Homo sapiens partial 2c-18m3 gene for immunoglobulin heavy chain
DEFINITION      variable region.
ACCESSION      AJ306240.1 GI:12641742
VERSION      2c-18m3 gene; immunoglobulin heavy chain; variable region.
KEYWORDS      human.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS      Weisgerber,K.
JOURNAL      Thesis (2000) Department of Medizinische Klinik,
              Universitaetskliniken des Saarlandes, Homburg, Germany
REFERENCE      2 (bases 1 to 246)
AUTHORS      Voswinkel,J., Weisgerber,K., Pfeundschnuh,M. and Gause,A.
TITLE      B lymphocyte involvement in ankylosing spondylitis: the heavy chain
              variable segment gene repertoire of B lymphocytes from germinal
              center-like foci in the synovial membrane indicates antigen
              selection
JOURNAL      Arthritis Res. 3 (3), 189-195 (2001)
MEDLINE      21199060
PUBMED      11299060
REFERENCE      3 (bases 1 to 246)
AUTHORS      Weisgerber,K.
TITLE      Direct Submission
JOURNAL      Submitted (06-JUN-2001) Weisgerber K., Universitaetskliniken des
              Saarlandes, Onkologisches Labor, 66424 Homburg, GERMANY
FEATURES
source      location/Qualifiers
              1..246
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /tissue_type="Synovial membrane"
              /rearranged
              1..246
              /gene="2c-18m3"
              <1..>246
              /gene="2c-18m3"
              /product="immunoglobulin heavy chain variable region"
V_region
BASE COUNT      65 a      59 c      68 g      54 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 246;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
        |||
Db      19 GACTACTACATGAGC 33

RESULT 7
AF297164      255 bp      DNA      linear      PRI 04-DEC-2000
LOCUS      Homo sapiens clone 5-4.3.10A immunoglobulin heavy chain variable
DEFINITION      region gene, partial cds.
ACCESSION      AF297164
VERSION      AF297164.1 GI:11528307
KEYWORDS      Homo sapiens.
SOURCE      Homo sapiens.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 255)
AUTHORS      Lee,J. and Lipsky,P.F.
TITLE      The V lambda J lambda repertoire in human fetal spleen: evidence
              for positive selection and extensive receptor editing
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 255)
AUTHORS      Lee,J. and Lipsky,P.F.
TITLE      Direct Submission
JOURNAL      Submitted (18-AUG-2000) Internal Medicine and Harold C. Simmons
              Arthritis Research Center, University of Texas Southwestern Medical
              Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA
FEATURES
source      location/Qualifiers
              1..255
              /organism="Homo sapiens"
              /isolate="f55"
              /db_xref="taxon:9606"
              /chromosome="14"
              /clone="5-4.3.10A"

```

MRNA
CDS
/cell_type="IgM+ B lymphocyte"
/tissue_type="spleen"
/dev_stage="fetus; 18 wk gestation"
/rearranged
<1..>255
/product="immunoglobulin heavy chain variable region"
<1..>255
/note="3-11"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="AA037215.1"
/db_xref="GI:11528308"
/translation="ASGFTESDYMSKIRQAPGKGLDWSYISDSSSTIYYADSVKGR
FTISRDNAKNSLYLQNSLRRAEDTAVYICARQDGPADITW"

BASE COUNT 63 a 61 c 72 g 59 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 255;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GACTACTACATGAGC 15
Db 22 GACTACTACATGAGC 36

RESULT 8
HSEM112V3 258 bp DNA linear PRI 27-MAR-1997
LOCUS H.sapiens rearranged Ig heavy chain variable region (EM2112VH3).
DEFINITION Y12032.1 GI:1914732
ACCESSION Y12032
VERSION Y12032.1
KEYWORDS V-region; immunoglobulin; immunoglobulin heavy chain; J-region;
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 258)
Foerster, N., Gellrich, S., Golembowski, S., Rutz, S., Audring, H.,
Sterry, W. and Jahn, S.
Analysis of VH genes rearranged in single B cells contained in
dermal infiltrates of patients with Mycosis fungoides
Unpublished
2 (bases 1 to 258)
Jahn, S.
Direct Submission
Submitted (21-MAR-1997) S. Jahn, University Hospital Charite,
Department of Dermatology, Schumannstrasse 20/21, 10117 Berlin, FRG

FEATURES
SOURCE
1..258
Location/Qualifiers
/organism="Homo sapiens"
/isolate="Patient EM"
/db_xref="taxon:9606"
/chromosome="14"
/cell_type="B lymphocyte"
/tissue_type="skin biopsy"
/dev_stage="adult"
/rearranged
/note="single B cell in a mycosis fungoides infiltrate"
1..258
/gene="EM2112VH3"
<1..>258
/gene="EM2112VH3"
/codon_start=1
/product="immunoglobulin heavy chain"
/protein_id="CAA72763.1"
/db_xref="GI:1914732"
/translation="SGFSFSDYMSKIRQAPGKGLDWSYISDSSSTIYYADSVKGR
FTISRDNAKNSLYLQNSLRRAEDTAVYICARQDGPADITW"

BASE COUNT 66 a 66 c 67 g 59 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GACTACTACATGAGC 15
Db 19 GACTACTACATGAGC 33

RESULT 9
HSA406644 261 bp DNA linear PRI 22-MAR-2001
LOCUS HSA406644
DEFINITION Hs sapiens partial VH3-11 gene for rearranged immunoglobulin
heavy chain variable region, isolate case4-PTGC1-171.
ACCESSION AJ406644
VERSION AJ406644.1 GI:10636712
KEYWORDS immunoglobulin heavy chain; variable region; VH3-11 gene.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 261)
Brauninger, A., Yang, W., Wacker, H., Rajewsky, K., Kuppers, R. and
Hansmann, M.L.
B-cell development in progressively transformed germinal centers:
similarities and differences compared with classical germinal
centers and lymphocyte-predominant Hodgkin disease
Blood 97 (3), 714-719 (2001)
21093792
PUBMED 1157489
2 (bases 1 to 261)
Brauninger, A.
Direct Submission
Submitted (21-SEP-2000) Brauninger, A., Pathology, University of
Frankfurt, Theodor Stern Kai 7, 60590, GERMANY
Location/Qualifiers
1..261
/organism="Homo sapiens"
/isolate="case4-PTGC1-171"
/db_xref="taxon:9606"
/rearranged
1..261
/gene="VH3-11"
<1..>261
/gene="VH3-11"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="CAC10708.1"
/db_xref="GI:10636713"
/translation="SGFTPSDYMSKIRQAPGKGLDWSYISDSSSTIYYADSVKGR
FTISRDNAKNSLYLQNSLRRAEDTAVYICARQDGPADITW"

FEATURES
SOURCE
1..261
Location/Qualifiers
/organism="Homo sapiens"
/isolate="case4-PTGC1-171"
/db_xref="taxon:9606"
/rearranged
1..261
/gene="VH3-11"
<1..>261
/gene="VH3-11"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="CAC10708.1"
/db_xref="GI:10636713"
/translation="SGFTPSDYMSKIRQAPGKGLDWSYISDSSSTIYYADSVKGR
FTISRDNAKNSLYLQNSLRRAEDTAVYICARQDGPADITW"

gene
CDS
1..261
/gene="VH3-11"
<1..>261
/gene="VH3-11"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="CAC10708.1"
/db_xref="GI:10636713"
/translation="SGFTPSDYMSKIRQAPGKGLDWSYISDSSSTIYYADSVKGR
FTISRDNAKNSLYLQNSLRRAEDTAVYICARQDGPADITW"

BASE COUNT 66 a 62 c 71 g 62 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GACTACTACATGAGC 15
Db 19 GACTACTACATGAGC 33

RESULT 10
HSA415895 273 bp DNA linear PRI 12-OCT-2001
LOCUS HSA415895
DEFINITION Hs sapiens partial IGWH3-8 gene for immunoglobulin heavy chain
variable region, donor MF, cell14.
ACCESSION AJ415895
VERSION AJ415895.1 GI:16117163

KEYWORDS IGVH gene; immunoglobulin heavy chain; variable region.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 273)
AUTHORS Brauninger, A., Goossens, T., Rajewsky, K. and Kuppers, R.
TITLE Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human
JOURNAL Eur. J. Immunol.
REFERENCE 2 (bases 1 to 273)
AUTHORS Goossens, T., Brauninger, A., Klein, U., Kuppers, R. and Rajewsky, K.
TITLE Receptor revision plays no major role in shaping the receptor repertoire of human memory B cells after the onset of somatic hypermutation
JOURNAL Eur. J. Immunol.
REFERENCE 3 (bases 1 to 273)
AUTHORS Brauninger, A.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2001) Brauninger A., Pathology, University of Frankfurt, Theodor Stern Kai 7, 60590, GERMANY
FEATURES
Source Location/Qualifiers
1..273
/organism="Homo sapiens"
/isolate="donor MF"
/db_xref="taxon:9606"
/clone="cell4"
/cell_type="B cell"
/rearranged
/note="naive lambda expressing"
1..273
/gene="IGVH3-8"
<1..>273
/gene="IGVH3-8"
/product="immunoglobulin heavy chain variable region"
BASE COUNT 72 a 61 c 78 g 62 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 273;
Best Local Similarity 100.0%; Pred. NO. 9.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15
Db 1 GACTACTACTGAGC 15
RESULT 11
LOCUS HS277321 278 bp DNA linear PRI 08-NOV-1996
DEFINITION H.sapiens gene for immunoglobulin heavy chain variable region, VH gene (V3-11).
ACCESSION 277321
VERSION 277321.1 GI:1465675
KEYWORDS Immunoglobulin heavy chain; immunoglobulin superfamily; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 278)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 278)
AUTHORS Kanzler, H., Kueppers, R., Hansmann, M.L. and Rajewsky, K.
TITLE Hodgkin and Reed/Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 278)
AUTHORS Kanzler, H.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-1996) Kanzler H., Institute for Genetics at the University of Cologne, Weyertal 121, Cologne 50931, FRG
REFERENCE 3 (bases 1 to 278)
AUTHORS Kanzler, H., Kuppers, R., Hansmann, M.L. and Rajewsky, K.

TITLE Hodgkin and Reed/Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells
JOURNAL J. Exp. Med. 184 (4), 1495-1505 (1996)
MEDLINE 97033523
PUBMED 8679220
FEATURES
Source Location/Qualifiers
1..278
/organism="Homo sapiens"
/isolate="patient 8"
/db_xref="taxon:9606"
/cell_type="Hodgkin-RS"
/tissue_type="Hodgkin's lymphoma"
/rearranged
1..278
/gene="VH (V3-11)"
1..278
/gene="VH (V3-11)"
/note="immunoglobulin heavy chain variable region"
BASE COUNT 69 a 64 c 82 g 63 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 278;
Best Local Similarity 100.0%; Pred. NO. 9.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15
Db 19 GACTACTACTGAGC 33
RESULT 12
LOCUS AF077461 279 bp DNA linear PRI 04-NOV-1998
DEFINITION Homo sapiens clone BF3N233E09 immunoglobulin heavy chain variable region gene, partial cds.
ACCESSION AF077461
VERSION AF077461.1 GI:3832595
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 279)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 279)
AUTHORS Brezinschek, H.P., Dörner, T., Farmer, N.L., Brezinschek, R.I. and Lipsky, P.E.
TITLE The influence of CD40-CD154 interactions on the Expressed Human Variable Heavy Chain Repertoire: Analysis of VH Genes Expressed by Individual B Cells of a Patient with X-Linked Hyper-Igm Syndrome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 279)
AUTHORS Brezinschek, H.P., Dörner, T., Farmer, N.L., Brezinschek, R.I. and Lipsky, P.E.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1998) Internal Medicine, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA
FEATURES
Source Location/Qualifiers
1..279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="BF3N233E09"
/cell_type="individual B cell"
/rearranged
/note="from patient with X-linked hyper-igm syndrome"
1..>279
/note="VH(D)JH"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="AAC70819.1"
/db_xref="GI:3832596"
/translation="VKRGSRLSCAAGFSDYVMIRAPGKLEWYSISSG
STIYVDSVKGRFTISRDNKNSLYLQMSLPAEDTAYTCARDIGVNG"

```

BASE COUNT      66 a      71 c      81 g      61 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 9, 7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GACTACTACATGAGC 15
        |||
Db      58 GACTACTACATGAGC 72

RESULT 13
AY013309      279 bp      DNA      linear      PRI 03-DEC-2001
LOCUS      Homo sapiens clone FMI-3.3.12A immunoglobulin heavy chain variable
DEFINITION      region gene, partial cds.
ACCESSION      AY013309.1 GI:17220530
VERSION
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE      Lee, J. and Lipsky, P.E.
JOURNAL      VH gene replacement in human fetal B cells: evidence for
REFERENCE      Immunoglobulin heavy chain receptor editing
AUTHORS      Unpublished
TITLE      2 (bases 1 to 279)
JOURNAL      Lee, J. and Lipsky, P.E.
AUTHORS      Direct Submission
TITLE      Submitted (08-NOV-2000) Internal Medicine and Harold C. Simmons
JOURNAL      Arthritis Research Center, University of Texas Southwestern Medical
FEATURES      Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA
SOURCE      Location/Qualifiers
1. 279
/organism="Homo sapiens"
/isolate="FMI"
/db_xref="taxon:9606"
/chromosome="14"
/clone="FMI-3.3.12A"
/cell_type="IgM+ B lymphocyte"
/tissue_type="bone marrow"
/dev_stage="fetal; 18 weeks gestation"
/rearranged
/note="donor VH3-8; recipient 3-11"
<1. >279
/product="immunoglobulin heavy chain variable region"
<1. >279
/note="CDR2 replacement"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="AA45480.1"
/db_xref="GI:17220531"
/translation="LRLSCAASGTFPSDYMSWIRQAPGKLEWYSISSSSSYTNY
ADVSGRFRTSRNNAKNSLYLQWNSLRADFTAVYVCARPHYSGSDYV"

BASE COUNT      70 a      67 c      77 g      65 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 9, 7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GACTACTACATGAGC 15
        |||
Db      40 GACTACTACATGAGC 54

RESULT 14
AY003824      283 bp      mRNA      linear      PRI 25-SEP-2001
LOCUS      Homo sapiens isolate sy-3M/9-H1 immunoglobulin mu heavy chain
DEFINITION

```

```

ACCESSION      variable region mRNA, partial cds.
VERSION      AY003824.1 GI:15777397
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE      1 (bases 1 to 283)
JOURNAL      Yavuz, S., Grammer, A.C., Yavuz, A.S., Nanki, T. and Lipsky, P.E.
REFERENCE      Comparative characteristics of mu chain and alpha chain transcripts
AUTHORS      expressed by individual tonsil plasma cells
JOURNAL      Mol. Immunol. 38 (1), 19-34 (2001)
MEDLINE      21376478
PUBMED      11483207
2 (bases 1 to 283)
Yavuz, S., Nanki, T., Yavuz, A.S. and Lipsky, P.E.
Direct Submission
Submitted (02-JUN-2000) Autoimmunity Branch, National Institute of
Arthritis and Musculoskeletal and Skin Diseases, 9000 Rockville Pike
Bldg. 9 IN124, Bethesda, MD 20892, USA
Location/Qualifiers
1. 283
/organism="Homo sapiens"
/isolate="sy-3M/9-H1"
/db_xref="taxon:9606"
/cell_type="plasma cell"
<1. >283
/note="VH3 family, 3-11/DP-35"
/codon_start=1
/product="immunoglobulin mu heavy chain variable region"
/protein_id="AAK11793.1"
/db_xref="GI:15777398"
/translation="SDYMSWIRQAPGKLEWYSISSGRTIYANSKGRFTSRD
NAKNSLYLQWNSLRADFTAVYVCATPYSGRNVLLLELVGQCTLVTVSSG"

BASE COUNT      74 a      72 c      76 g      61 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 283;
Best Local Similarity 100.0%; Pred. No. 9, 7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GACTACTACATGAGC 15
        |||
Db      4 GACTACTACATGAGC 18

RESULT 15
HMMIGL2E5      291 bp      mRNA      linear      PRI 06-JAN-1995
LOCUS      Human (cell line: L2E5) Ig rearranged H-chain mRNA V-region,
DEFINITION      partial cds.
ACCESSION      L04325.1 GI:186055
VERSION      L04325.1
KEYWORDS      V-region; immunoglobulin heavy chain V-region.
SOURCE      Homo sapiens fetus liver cDNA to mRNA.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE      1 (bases 1 to 291)
JOURNAL      Hillson, J.L., Opplinger, I.R., Sasso, E.H., Milner, E. and Wener, M.H.
REFERENCE      The emerging B cell repertoire: Influence of developmental stage
AUTHORS      and interindividual variation
JOURNAL      J. Immunology 149, 3741-3752 (1992)
2 (sites)
Hillson, J.L., Karf, N.S., Opplinger, I.R., Mannik, M. and Sasso, E.H.
The structural basis of germline-encoded VH3 immunoglobulin binding
to staphylococcal protein A
J. Exp. Med. 178 (1), 331-336 (1993)
MEDLINE      93301610
PUBMED      8315388
JOURNAL      Location/Qualifiers
1. 291

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14q32.33"
/cell_line="L2E5"
/cell_type="B-lymphocyte"
/tissue_type="liver"
/dev_stage="fetus"
1..291
/gene="IGHV@"
<1..>291
/gene="IGHV@"
/note="G00-128-528"
1..>291
/gene="IGHV@"
/note="This CDS feature is included to show the
translation of the corresponding V_region. Presently
translation qualifiers on V_region features are illegal."
/codon_start=1
/protein_id="AA59103.1"
/db_xref="GI:619626"
/translation="GGLVKGPGSLRLSCAAGFTESDYMSWIRQAPKGLQVSYIS
SSSYTYADSVKGRFTISRDNKNSLYLQMSLRADFTAVYCARHWYFDL"
1..291
/partial
/gene="IGHV@"
/product="immunoglobulin heavy chain V-region"
/note="G00-128-528"
1..66
/gene="IGHV@"
/note="Framework 1; G00-128-528"
67..87
/gene="IGHV@"
/note="CDR-1; G00-128-528"
88..123
/gene="IGHV@"
/note="Framework 2; G00-128-528"
124..174
/gene="IGHV@"
/note="CDR-2; G00-128-528"
175..270
/gene="IGHV@"
/note="Framework 3; G00-128-528"
271..291
/gene="IGHV@"
/note="CDR-3; G00-128-528"
71 a 74 c 80 g 66 t
BASE COUNT
ORIGIN

misc_feature
/note="IGHV@"
/note="Framework 1; G00-128-528"
67..87
/gene="IGHV@"
/note="CDR-1; G00-128-528"
88..123
/gene="IGHV@"
/note="Framework 2; G00-128-528"
124..174
/gene="IGHV@"
/note="CDR-2; G00-128-528"
175..270
/gene="IGHV@"
/note="Framework 3; G00-128-528"
271..291
/gene="IGHV@"
/note="CDR-3; G00-128-528"
71 a 74 c 80 g 66 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 291;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY .1 GACTACTACATGAGC 15
|||||
Db 67 GACTACTACATGAGC 81

```

Search completed: June 26, 2003, 03:47:29
Job time : 427.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-13

Sequence: 1 gactactacatgac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002: *
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	20	AA332393
2	15	100.0	294	21	AA246860
3	15	100.0	329	24	AB894229
4	15	100.0	348	22	AA503478
5	15	100.0	351	24	AA43586
6	15	100.0	354	22	AA503470
7	15	100.0	366	20	AA231651
8	15	100.0	369	22	AA503533
9	15	100.0	715	16	AA078949

10	14.2	94.7	379	21	AA000031	Human secreted pro
11	14	93.3	363	21	AA503474	DNA encoding anti-
12	14	93.3	606	16	AA078960	Human immunoglobul
13	14	93.3	1422	24	ABK65583	Rat sequence diffe
14	14	93.3	2751	24	AAD31625	Arabidopsis thalia
15	14	93.3	3236	24	AAD31644	Arabidopsis thalia
16	14	93.3	8002	24	AAD31629	Arabidopsis thalia
17	13.4	89.3	21	22	AA507200	Rt-PCR down-primer
18	13.4	89.3	228	24	ABN26925	Human ORFX polyomc
19	13.4	89.3	348	22	AA503446	DNA encoding anti-
20	13.4	89.3	349	22	AA503504	DNA encoding anti-
21	13.4	89.3	365	22	AAK83680	Human immune/haema
22	13.4	89.3	366	22	AA503447	DNA encoding anti-
23	13.4	89.3	366	22	AA503526	DNA encoding anti-
24	13.4	89.3	368	22	AAK83677	Human immune/haema
25	13.4	89.3	372	18	AAK58268	Lead binding Mab 2
26	13.4	89.3	372	18	AA503448	DNA encoding anti-
27	13.4	89.3	467	22	ABA51889	Human foetal liver
28	13.4	89.3	467	22	ABA21709	Probe #175 for gen
29	13.4	89.3	467	22	AAK00186	Human brain expres
30	13.4	89.3	467	22	AAK25626	Human bone marrow
31	13.4	89.3	467	22	AAI10253	Probe #186 for gen
32	13.4	89.3	467	22	AAI10253	Probe #186 used to
33	13.4	89.3	467	22	AAI00190	Human genome-deriv
34	13.4	89.3	467	22	AB500195	Human secreted pro
35	13.4	89.3	529	21	AAK08128	Aspergillus oryzae
36	13.4	89.3	674	21	AAE13480	Aspergillus oryzae
37	13.4	89.3	675	21	AAE12980	Trichoderma reesei
38	13.4	89.3	729	21	AAE15002	p4-2 single chain
39	13.4	89.3	777	24	AA597135	Artificial synthe
40	13.4	89.3	1080	20	AAZ21160	Artificial synthe
41	13.4	89.3	1086	20	AAZ21157	Artificial synthe
42	13.4	89.3	1113	20	AAZ21159	Novel human coding
43	13.4	89.3	1167	24	ABN59693	Morine Kv6.2 cDNA.
44	13.4	89.3	1466	20	AAZ22804	Human cancer assoc
45	13.4	89.3	1571	21	AACT7699	

ALIGNMENTS

RESULT 1	AA332393	standard; DNA; 15 BP.
ID	AA332393	
AC	AA332393	
XX		
DT	17-JUN-1999	(first entry)
XX		
DE	AB2 variable heavy (VH) chain CDRL encoding DNA.	
XX		
KW	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;	
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;	
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;	
KW	myelodysplasia; myelotoxic chemotherapy; leukaemia; tumour; MusK; CDR;	
KW	neuromuscular; muscular dystrophy; complementarity determining region;	
KW	variable heavy chain; variable light chain; VH; VL; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9910494-A2.	
XX		
PD	04-MAR-1999.	
XX		
PF	21-AUG-1998; 98WO-US17364.	
XX		
PR	25-AUG-1997; 97US-0918148.	
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;	
XX		
DR	WPI; 1999-204666/17.	

DR P-PSDB: AAY06693.
 XX
 PT New thrombopoietin receptor agonist antibodies - useful for
 PT treating immunological or hematological disorders
 XX
 PS Claim 10; Page 76; 86pp; English.

XX The invention relates to an agonist antibody (Ab) which binds to a
 CC thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
 CC be used in the same way and for the same indications as thrombopoietin
 CC (TPO). They can stimulate proliferation, differentiation or growth of
 CC megakaryocytes. They may also be able to stimulate megakaryocytes to
 CC increase platelet production. They can be used for treating
 CC immunological or hematopoietic disorders, especially thrombocytopenia.
 CC Thrombocytopenia - associated bone marrow hypoplasia (e.g. aplastic anemia
 CC following chemotherapy or bone marrow transplant) may be effectively
 CC treated with the antibody compounds as well as disorders such as
 CC disseminated intravascular coagulation (DIC), immune thrombocytopenia
 CC (HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
 CC congenital thrombocytopenia, thrombotic thrombocytopenia and
 CC myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
 CC treatment of solid tumours or leukaemia, myeloblastic chemotherapy for
 CC autologous or allogeneic bone marrow transplant, myelodysplasia,
 CC idiopathic aplastic anemia, congenital thrombocytopenia, and immune
 CC thrombocytopenia. The antibodies which bind to the TPO-R can be
 CC used for improving neuromuscular function in a patient, e.g. in muscular
 CC dystrophy. The products can also be used for detection and diagnosis. The
 CC antibodies have a longer half-life than the natural ligand for the TPO-R.
 CC Sequences AAX32387-X32413 represent DNA fragments encoding for the CDRI,
 CC CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
 CC (VL) chains of antibodies Ab1 to Ab6.

XX Sequence 15 BP; 5 A; 4 C; 3 G; 3 T; 0 other;

Query Match 100.0%; Score 15; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. NO. 52;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GACTACTACATGAGC 15
 Db 1 GACTACTACATGAGC 15

RESULT 2

AAZ46860
 ID AAZ46860 standard; DNA: 294 BP.

XX AAZ46860;

DT 10-APR-2000 (first entry)

DE Human immunoglobulin IGHV3-11 Vh gene sequence.

XX Anti-xenogeneic antigen; antibody; xenograft rejection; diabetes;
 KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
 KW anti-amyloid antibody; Alzheimer's disease; or anti-insulin; ss.

OS Homo sapiens.

XX W0200185797-A1.

XX 18-NOV-1999.

XX 07-MAY-1999; 99WO-US09971.

XX 08-MAY-1998; 9805-0084725.

XX (GRAM/) CRAMER D V.
 PA (KEAR/) KEARNS-JONKER M.

XX Cramer DV, Kearns-Jonker M;
 DR WPI; 2000-116276/10.

DR P-PSDB: AAY56862.
 XX
 PT Use of antisense oligonucleotides for preventing xenograft rejection or
 PT treating autoimmune conditions, e.g. systemic lupus erythematosus,
 PT rheumatoid arthritis, Alzheimer's disease or diabetes
 XX
 PS Claim 11; Page 16; 34pp; English.

XX The invention provides new methods for preventing expression of anti-
 CC xenogeneic antigen antibodies. The method comprises (i) selecting genes,
 CC mRNA, or proteins associated with the expression or regulation of
 CC antibodies to xenogeneic tissue, organ or antigens; (ii) obtaining ONS
 CC which are anti-sense to a target selected from genes, mRNAs and fragments
 CC which are associated with expression or regulation of expression of the
 CC antibodies or aromatic amino acid oligomers comprising an amino acid
 CC sequence selected from antibody fragments targeted to minor microgrooves
 CC of the polynucleotides encoding them having at least one aromatic amino
 CC acid, and (iii) administering the ONS obtained to a subject to bind to
 CC its target and prevent, reduce or inhibit expression of anti-xenogeneic
 CC antigen antibody. The methods can be used for preventing, reducing or
 CC inhibiting xenograft rejection. The xenograft may be e.g. liver, heart
 CC and kidney organs, or islets of Langerhans tissues. The methods can also
 CC be used for treating autoimmune diseases or conditions e.g. systemic
 CC lupus erythematosus, rheumatoid arthritis, anti-amyloid antibodies
 CC (Alzheimer's disease) or anti-insulin auto-antibodies (diabetes). The
 CC products can also be used for detection of anti-xenogeneic antigen
 CC antibodies and nucleic acids encoding them. The present sequence
 CC represents a human immunoglobulin IGHV3-11 Vh gene sequence binding to
 CC target nucleic acids.

XX Sequence 294 BP; 68 A; 72 C; 90 G; 64 T; 0 other;

Query Match 100.0%; Score 15; DB 21; Length 294;
 Best Local Similarity 100.0%; Pred. NO. 65;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GACTACTACATGAGC 15
 Db 91 GACTACTACATGAGC 105

RESULT 3

ABA94229
 ID ABA94229 standard; DNA: 329 BP.

XX ABA94229;

DT 13-MAR-2002 (first entry)

DE CB2IE7 heavy chain variable region nucleotide sequence.

XX Neuroendocrine; central nervous system; CNS; SH1GM22; LYM 22; AKJRA;
 KW ebyH1GM MS119D10; eby H1GM CB2BG8; CB2IE12; CB2IE7; MS119E5; v1rucide;
 KW antiparkinsonian; neuroprotective; nootropic; vulnerable; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..329

XX /transl_except= "(pos: 1..2; aa: Xaa)"
 XX /note= "Xaa = unknown"

XX W0200185797-A1.

XX 15-NOV-2001.

XX 30-MAY-2000; 2000WO-US14902.

XX 10-MAY-2000; 2000US-0568351.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION RES.
 XX

PI Rodriguez M, Miller DJ, Pease LR;
 XX WPI: 2002-066596/09.
 DR P-PSDB: ABB07182.
 XX
 PT Novel neuromodulatory agent (a human IGM monoclonal antibody),
 PT promoting neurite outgrowth, regeneration, remyelination and
 PT neuroprotection in central nervous system, useful to treat
 PT post-infectious encephalomyelitis -
 PS Claim 43; Fig 41; 219pp; English.
 XX
 XX The invention provides a neuromodulatory agent (I) capable of promoting
 CC neurite outgrowth, regeneration, remyelination and neuroprotection in
 CC central nervous system (CNS). (I) is capable of inducing remyelination,
 CC promoting cellular proliferation of glial cells, and promoting Ca2+
 CC signaling with oligodendrocytes. An humanised antibody to (I) can be
 CC selected from antibody SH19M22 (LYM 22), ebvH19M MS119D10, ebv H19M
 CC CB2B68, AKR4, CB21E12, CB21E7 or MS119E5. (I) is useful for stimulating
 CC remyelination of CNS axons, stimulating proliferation of glial cells in
 CC CNS axons, or treating demyelinating disease of CNS in a mammal in need
 CC of such therapy. (I) is capable of binding to structures and cells within
 CC CNS. (I) is preferably useful for treating a demyelinating disease of CNS
 CC of a mouse infected with Strain DA of Theiler's murine encephalomyelitis
 CC (TMEV) or for treating a human being having multiple sclerosis, or a post-
 CC human or domestic animal with a viral demyelinating disease, or a post-
 CC neural disease of CNS. (I) is also useful for an in vitro method of
 CC stimulating the proliferation of glial cells from mixed cell culture.
 CC (I) is also useful for stimulating remyelination of CNS axons. The
 CC antibodies are useful for preventing infection by a bacterium, virus or
 CC like pathogen that causes demyelination or other neurodegenerative
 CC condition in a subject. Methods where (I) is administered to a patient
 CC are useful for treating multiple sclerosis, Parkinson's disease,
 CC Alzheimer's disease, amotrophic lateral sclerosis (ALS), a viral
 CC demyelinating disease, CNS diseases, and other conditions in the CNS
 CC where nerves are damaged as by trauma. The present sequence represents
 CC the CB21E7 heavy chain variable region nucleotide sequence.
 CC
 SO Sequence 329 BP; 81 A; 81 C; 96 G; 71 T; 0 other;
 Query Match 100.0%; Score 15; DB 24; Length 329;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACTACTACATGAGC 15
 ||||||||||||||||
 DB 66 GACTACTACATGAGC 80
 RESULT 4
 AAS03478
 ID AAS03478 standard; cDNA; 348 BP.
 XX AAS03478;
 AC
 XX 29-AUG-2001 (first entry)
 DT
 XX
 DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 65.
 XX
 KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 KW heart disease; complementarity determining region; CDR; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200127279-A1.
 PD 19-APR-2001.
 XX
 PF 11-OCT-2000; 2000MO-GB03900.
 XX
 PR 12-OCT-1999; 99US-0158812.
 XX
 PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.

XX
 PI Edwards BM, Main SH, Vaughan TJ;
 XX WPI: 2001-282031/29.
 DR P-PSDB: AAU02578.
 XX
 PT Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -
 PS Disclosure: Page 141; 182pp; English.
 XX
 XX AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.
 CC
 SO Sequence 348 BP; 79 A; 91 C; 108 G; 70 T; 0 other;
 Query Match 100.0%; Score 15; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACTACTACATGAGC 15
 ||||||||||||||||
 DB 91 GACTACTACATGAGC 105
 RESULT 5
 AAL43586
 ID AAL43586 standard; DNA; 351 BP.
 XX AAL43586;
 AC
 XX 05-SEP-2002 (first entry)
 DT
 XX
 DE Dig3 antibody heavy chain coding sequence.
 XX
 XX 26-10 scFv antibody light chain; complementarity determining region 3;
 KW CDR3; binding protein production; catalytic protein production; ds;
 KW ligand structure determination; antibody isolation; dig3 heavy chain.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..351
 FT /*tag= a
 FT /*partial
 FT /*product= "Dig3 heavy chain"
 FT /*note= "No start or stop codon is given"
 XX
 PN WO200234886-A2.
 PD 02-MAY-2002.
 XX
 PF 26-OCT-2001; 2001MO-US46795.
 XX
 PR 27-OCT-2000; 2000US-0699023.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.

PI Chen G, Hayhurst A, Thomas JC, Iverson BL, Georgiou G;
 XX WPI: 2002-500078/53.
 DR P-PSDB; AA015187.
 XX
 PT Obtaining bacterium having nucleic acid encoding binding protein that
 PT binds target ligand, or a nucleic acid encoding catalytic protein that
 PT catalyzes reaction involving target substrate by display-less library
 PT screening -
 XX
 PS Example 4; Page 97; 98pp; English.
 XX
 CC The invention comprises a method of obtaining a bacterium containing a
 CC nucleic acid encoding a binding protein capable of binding a target
 CC ligand, or a nucleic acid encoding a catalytic protein that catalyses a
 CC reaction involving a target substrate. The method of the invention allows
 CC the isolation of a cell expressing a nucleic acid encoding a binding
 CC protein capable of binding a target ligand, and consequently the
 CC isolation of the nucleic acid. The binding protein produced by the method
 CC of the invention is useful in determining the structure of a target
 CC ligand. The method of the invention is useful for the de novo isolation
 CC of antibodies from large repertoire libraries and for monitoring
 CC production during protein manufacturing. The method can also be used for
 CC monitoring production of a particular byproduct of a biological reaction.
 CC The present DNA sequence encodes the dig3 scfv antibody heavy chain which
 CC is used in an example of the invention.
 XX
 S0 Sequence 351 BP; 77 A; 85 C; 109 G; 80 T; 0 other;
 XX
 Query Match 100.0%; Score 15; DB 24; Length 351;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACTACTACATGAGC 15
 |||||||||||||
 Db 91 GACTACTACATGAGC 105
 XX
 RESULT 6
 AAS03470
 ID AAS03470 standard; cDNA; 354 BP.
 XX
 AC AAS03470;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 58.
 XX
 KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 KW heart disease; complementarity determining region; CDR; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200127279-A1.
 PN
 PD 19-APR-2001.
 PD
 XX
 PF 11-OCT-2000; 2000MO-GB03900.
 PF
 XX
 PR 12-OCT-1999; 99US-0158812.
 PR
 XX
 PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 PA
 PI Edwards BM, Main SH, Vaughan TJ;
 PI
 DR WPI: 2001-282031/29.
 DR
 XX P-PSDB; AAU02570.
 XX
 PT Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -
 XX
 CC Disclosure; Page 136; 182pp; English.
 PS

XX
 CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.
 XX
 S0 Sequence 354 BP; 83 A; 84 C; 108 G; 79 T; 0 other;
 XX
 Query Match 100.0%; Score 15; DB 22; Length 354;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACTACTACATGAGC 15
 |||||||||||||
 Db 91 GACTACTACATGAGC 105
 XX
 RESULT 7
 AA231651
 ID AA231651 standard; DNA; 366 BP.
 XX
 AC AA231651;
 XX
 DT 13-JAN-2000 (first entry)
 DT
 XX
 DE Coding sequence for VH domain CDR of anti-estradiol antibody.
 DE
 XX
 KW Estradiol; complementarity determining region; CDR; estriol-3-sulphate;
 KW antibody antigen binding domain; steroid hormone; estriol; testosterone;
 KW dihydrotestosterone; progesterone; estriol 3-beta-di-glucuronide;
 KW menstrual cycle; hormone replacement therapy; oestrogen secreting tumour;
 KW diagnosis; VH domain; ds.
 XX
 OS Homo sapiens.
 OS
 XX
 PN US5977319-A.
 PN
 XX
 PD 02-NOV-1999.
 PD
 XX
 PF 21-OCT-1997; 97US-0958201.
 PF
 XX
 PR 21-OCT-1996; 96US-0028897.
 PR
 XX
 PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 PA
 PI Johnson KS, Pope AR, Pritchard K, Williams AJ;
 PI
 DR WPI: 1999-619713/53.
 DR
 XX P-PSDB; AAY43253.
 XX
 PT New specific binding partners for estradiol, used for monitoring
 PT estradiol levels during the menstrual cycle, in hormone replacement
 PT therapy and for diagnosing oestrogen secreting tumours -
 PT
 XX
 PS Disclosure; Column 21-22; 26pp; English.
 PS
 CC This sequence encodes a VH domain complementarity determining region
 CC (CDR) from an antibody specific for estradiol. The invention relates to
 CC specific binding members (sbp) comprising a polypeptide that comprises an
 CC antibody antigen binding domain (AABD) which has a dissociation constant
 CC of less than 1.0x10⁻⁸M for estradiol, and a dissociation constant of at

CC least 500-fold higher for the steroid hormones selected from estradiol, testosterone, dihydrotestosterone, progesterone, estradiol-3-sulphate and estradiol-3-beta-di-glucuronide, where the polypeptide comprises an antibody VH domain. The sbps can be used in an immunoassay for determining the presence or absence of estradiol in a sample. They can be used for monitoring estradiol levels, e.g. during the menstrual cycle, in hormone replacement therapy and for diagnosing oestrogen secreting tumours. The sbps can provide for discrimination between estradiol and other related steroids.

XX Sequence 366 BP; 80 A; 98 C; 110 G; 77 T; 1 other;

Query Match 100.0%; Score 15; DB 20; Length 366;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
DB 91 GACTACTACATGAGC 105

RESULT 8

AA03533 standard; CDNA; 369 BP.

AA03533;

29-AUG-2001 (first entry)

DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 116.

Antibody: adipocyte; heavy chain; light chain; obesity; FAT; heart disease; complementarity determining region; CDR; ss.

Homo sapiens.

WO200127279-A1.

19-APR-2001.

11-OCT-2000; 2000WO-GB03900.

12-OCT-1999; 99US-0158812.

(CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.

Edwards BM, Main SH, Vaughan TJ;

WPI: 2001-282031/29.

P-PSDB: AAU02633.

Panel of specific binding members of antibody molecules which bind to whole adipocytes is used in the treatment of obesity and obesity related diseases -

PS Disclosure: Page 175; 182pp: English.

XX AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
CC and light chain coding sequences of the invention. The antibodies can be
CC used in the treatment of obesity and obesity related diseases. The
CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
CC mass of an obese patient or the antibody can be used as a therapeutic
CC itself. Antibodies binding specifically to adipocytes can be used to
CC activate the immune system to destroy the cells by complement mediated
CC lysis. The antibodies may be labeled with a detectable label such as
CC radiolabel, fluorescent or chemical group and used in methods of
CC diagnosis in human subjects e.g. to determine the presence of adipocyte
CC antigen on the surface of an adipocyte to detect or determine the
CC presence or level of adipocytes in a cell or tissue sample. The
CC antibodies can be used as an alternative means of treatment for obese
CC patients other than undergoing surgery to remove excess fat. Antibodies
CC for different types of fat deposits can also be produced e.g. intra
CC abdominal fat associated with heart disease.

XX Sequence 369 BP; 81 A; 90 C; 119 G; 79 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 369;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
DB 91 GACTACTACATGAGC 105

RESULT 9

AA078949 standard; DNA; 715 BP.

AA078949;

01-AUG-1995 (first entry)

Human immunoglobulin Vh gene #11.

XX Primer; PCR; amplify: human; immunoglobulin: variable; heavy chain;
KW cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 206..653

FT /*tag= a /product= human immunoglobulin variable heavy chain

FT Intron 252..348

FT /*tag= b 654..656

FT /*tag= c /note= "miscellaneous signal, does not conform to

terminator or splice site sequence"

XX WO9426895-A.

XX 24-NOV-1994.

XX 10-MAY-1993; 93WO-JP00603.

XX 10-MAY-1993; 93WO-JP00603.

XX (NIBS) JAPAN TOBACCO INC.

XX Honjo T, Matsuda F;

XX WPI: 1995-006791/01.

XX P-PSDB: AAR66305.

XX DNA fragment comprising human immunoglobulin Vh genes - for the

XX production of human immunoglobulin in mammalian hosts

XX Claim 20: Page 44-45; 130pp: Japanese.

XX A series of genes (AA078939-79002) encoding human immunoglobulin
CC variable heavy chains. The genes were isolated and cloned from a series
CC of cosmid constructs: Y202, Y103, Y21, Y6, Y24, 3-31, M84, M118 and M131,
CC by PCR amplification using primers AA078917-38. The genes are subdivided
CC into 5 families of Vh genes. The fragments cover a region of 800 kb.
CC The DNA fragments were isolated from high molecular weight DNA from
CC human placenta. The DNA was partially digested with RsaI restriction
CC enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
CC fragments were collected. The fragments were ligated with ClaI-digested
CC cosmid vector pJB81. The ligation products were in vitro packed and
CC infected into E.coli 490A. The fragments were then subcloned by colony
CC hybridisation. The Vh genes and the DNA fragments encoding them are
CC useful in producing human immunoglobulin in mammalian hosts.

XX Sequence 715 BP; 170 A; 162 C; 209 G; 174 T; 0 other;

Query Match 100.0%; Score 15; DB 16; Length 715;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
 |||||
 DB 450 GACTACTACATGAGC 464

RESULT 10

AAC00031
 ID AAC00031 standard; cDNA; 379 BP.

AC AAC00031;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 29.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EPI033401-A2.

PD 06-SEP-2000.

PE 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

DR P-PSDB; AAG00025.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 29; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SQ Sequence 379 BP; 83 A; 88 C; 102 G; 85 T; 21 other;

Query Match 94.7%; Score 14.2; DB 21; Length 379;
 Best Local Similarity 86.7%; Pred. No. 1.9e+02;

Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
 |||||

DB 227 GACTACTACATGAGC 241

RESULT 11

AAS03474
 ID AAS03474 standard; cDNA; 363 BP.

AC AAS03474;

DT 29-AUG-2001 (first entry)

DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 62.

KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 heart disease; complementarity determining region; CDR; ss.

OS Homo sapiens.

PN WO200127279-A1.

PD 19-APR-2001.

PE 11-OCT-2000; 2000WO-GB03900.

PR 12-OCT-1999; 99US-0158812.

PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.

PI Edwards BM, Main SH, Vaughan TJ;

DR WPI; 2001-282031/29.

DR P-PSDB; AA002574.

XX Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -

PS Disclosure; Page 139; 182pp; English.

XX AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.

SQ Sequence 363 BP; 84 A; 91 C; 103 G; 85 T; 0 other;

Query Match 93.3%; Score 14; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 14
 |||||

DB 91 GACTACTACATGAGC 104

RESULT 12

AA078960
 ID AA078960 standard; DNA; 606 BP.

AC AA078960;

DT 03-AUG-1995 (first entry)

DE Human immunoglobulin Vh gene #22.

KW Primer: PCR; amplify: human; immunoglobulin; variable; heavy chain;
 KM cosmid; placenta; vector; pUB81; E.coli; mammalian; ds.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 85..544
 FT /tag= a
 FT /product= human immunoglobulin variable heavy chain
 FT Intron 131..233
 FT /tag= b
 FT misc_signal 428..430
 FT /tag= c
 FT /note= "non-functional termination codon"
 FT misc_signal 545..547
 FT /tag= d
 FT /note= "miscellaneous signal, does not conform to
 termination or splice site sequence"
 XX
 PN WO9426895-A.
 XX
 PD 24-NOV-1994.
 XX
 PF 10-MAY-1993; 93WO-JP00603.
 XX
 PR 10-MAY-1993; 93WO-JP00603.
 XX
 PA (NISR) JAPAN TOBACCO INC.
 XX
 PI Honjo T, Matsuda F;
 XX
 DR WPI; 1995-006791/01.
 DR P-PSDB; AAR66314.
 XX
 PT DNA fragment comprising human immunoglobulin Vh genes - for the
 PT production of human immunoglobulin in mammalian hosts
 XX
 PS Claim 31; Page 59-60; 130pp; Japanese.
 XX
 CC A series of genes (AA078939-79002) encoding human immunoglobulin
 CC variable heavy chains. The genes were isolated and cloned from a series
 CC of cosmid constructs: Y202, Y103, Y21; Y6:Y24; 3-31; M84; M18 and M131,
 CC by PCR amplification using primers AA078917-38. The genes are subdivided
 CC into 5 families of Vh genes. The fragments cover a region of 800 Kb.
 CC The DNA fragments were isolated from high molecular weight DNA from
 CC human placenta. The DNA was partially digested with Taqi restriction
 CC enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
 CC fractions were collected. The fragments were ligated with ClaI-digested
 CC cosmid vector pUB81. The ligation products were in vitro packed and
 CC infected into E.coli 490A. The fragments were then subcloned by colony
 CC hybridisation. The Vh genes and the DNA fragments encoding them are
 CC useful in producing human immunoglobulin in mammalian hosts.
 XX
 SQ Sequence 606 BP; 150 A; 128 C; 184 G; 144 T; 0 other;
 XX
 Query Match 93.3%; Score 14; DB 16; Length 606;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ACTACTACATGAGC 15
 DB 336 ACTACTACATGAGC 349
 XX
 RESULT 13
 ID ABR63583 standard; cDNA; 1422 BP.
 XX
 AC ABR63583;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Rat sequence differentially expressed in response to a hepatotoxin #1490.

XX
 KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KM differential expression; centrilobular necrosis; steatosis.
 OS Rattus norvegicus.
 XX
 PN WO200210453-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-US23872.
 XX
 PR 31-JUL-2000; 2000US-222040P.
 PR 02-NOV-2000; 2000US-244880P.
 PR 11-MAY-2001; 2001US-290029P.
 PR 15-MAY-2001; 2001US-290645P.
 PR 22-MAY-2001; 2001US-292336P.
 PR 06-JUN-2001; 2001US-295798P.
 PR 13-JUN-2001; 2001US-297457P.
 PR 19-JUN-2001; 2001US-298884P.
 PR 09-JUL-2001; 2001US-303459P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
 XX
 DR WPI; 2002-241625/29.
 XX
 PT Predicting toxic effects of compounds or the progression of these toxic
 PT effects by determining the changes in gene expression in tissues or
 PT cells exposed to the toxin and comparing these to gene expression in
 PT unexposed tissues or cells -
 XX
 PS Claim 1; Seq ID No 1490; 239pp; English.
 XX
 CC The invention relates to methods for predicting toxic effects of
 CC compounds or the progression of these toxic effects by determining the
 CC global changes in gene expression in tissues or cells exposed to the
 CC toxin and comparing these to gene expression in unexposed tissues or
 CC cells. Also included are methods of predicting at least one toxic
 CC effect of a compound or progression of a toxic effect, preferably the
 CC hepatotoxicity of a compound, comprising detecting the level of
 CC expression in a tissue or cell sample exposed to the compound of two or
 CC more genes listed in the specification, where differential expression of
 CC the genes is indicative of at least one toxic effect or progression.
 CC The method can also be used to identify an agent which modulates the
 CC toxic response and predict cellular pathways that a compound modulates
 CC in a cell. The methods utilise a set of at least two probes (on a solid
 CC support in kit form), where each of the probes comprises a sequence that
 CC specifically hybridises to a gene listed in the specification, a computer
 CC system comprising a database containing information identifying the
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
 CC set of genes comprising at least two genes listed in the specification,
 CC and a user interface to view the information used to present information
 CC identifying the expression level in a tissue or cell of at least one gene
 CC listed in the specification. The method is useful for elucidating global
 CC changes in gene expression and for identifying toxicity markers in
 CC tissues or cell exposed to a known toxin. The genes may be used as
 CC toxicity markers in drug screening and toxicity assays. The genes and
 CC gene expression information may be used as diagnostic markers for the
 CC prediction or identification of the physiological state of tissue or cell
 CC sample that has been exposed to a compound or agent. Hepatotoxicity
 CC is characterised by centrilobular necrosis and steatosis. The present
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
 CC which is differentially expressed in response to a hepatotoxic agent.
 XX
 SQ Sequence 1422 BP; 425 A; 305 C; 246 G; 446 T; 0 other;
 XX
 Query Match 93.3%; Score 14; DB 24; Length 1422;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GACTACTACATGAG 14

Db 347 GACTACTACATGAGC 360

RESULT 14
AAD31625

ID AAD31625 standard; cDNA: 2751 BP.

XX AAD31625;

DT 18-JUN-2002 (first entry)

DE Arabidopsis thaliana GT1209 gene cDNA coding sequence.

KW GT1209 gene; herbicide; plant tissue; cell; seedling growth; development;

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 1..2751

FT /tag= a

PN WO200212273-A2.

PD 14-FEB-2002.

PF 01-AUG-2001; 2001WO-EP08910.

PR 03-AUG-2000; 2000US-222779P.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Levin JZ, Megrich Glover L, Budziszewski GJ;

DR WPI: 2002-241730/29.

DR P-PSDB: AAE19876.

PT New polypeptide having GT1209, GT1354 or GT0946 activity, obtained from

PT Arabidopsis, useful as herbicide targets in screening assays to

PT identify the inhibitors or potential herbicides

PS Claim 2: Page 60-64; 82pp; English.

XX The invention relates to a polypeptide having GT1209, GT1354 or GT0946

CC activity obtained from Arabidopsis. The invention may also be applied to

CC the development of herbicide tolerant plants, plant tissues, plant seeds

CC and plant cells. The polypeptide is useful as herbicide targets in

CC screening assays to identify potential herbicides and inhibitors of

CC GT1209, GT1354 or GT0946 activity. A compound having herbicidal activity

CC is useful for suppressing the growth of a plant. The newly discovered

CC GT1209, GT1354 or GT0946 genes are essential for seedling growth and

CC development. The present sequence is Arabidopsis thaliana GT1209 gene

CC cDNA coding sequence.

XX Sequence 2751 BP; 732 A; 595 C; 603 G; 821 T; 0 other;

QY Query Match 93.3%; Score 14; DB 24; Length 2751;

Best Local Similarity 100.0%; Pred. No. 2.8e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1998 ACTACTACATGAGC 2011

RESULT 15

AAD31644

ID AAD31644 standard; cDNA: 3236 BP.

XX AAD31644;

AC

XX

DT 18-JUN-2002 (first entry)

DE Arabidopsis thaliana GT1209 gene cDNA second coding sequence.

KW GT1209 gene; herbicide; plant tissue; cell; seedling growth; development;

SS.

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 1..1056

FT /tag= a

PN WO200212273-A2.

PD 14-FEB-2002.

PF 01-AUG-2001; 2001WO-EP08910.

PR 03-AUG-2000; 2000US-222779P.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Levin JZ, Megrich Glover L, Budziszewski GJ;

DR WPI: 2002-241730/29.

DR P-PSDB: AAE19879.

PT New polypeptide having GT1209, GT1354 or GT0946 activity, obtained from

PT Arabidopsis, useful as herbicide targets in screening assays to

PT identify the inhibitors or potential herbicides

PS Example 3; Page 78-81; 82pp; English.

XX The invention relates to a polypeptide having GT1209, GT1354 or GT0946

CC activity obtained from Arabidopsis. The invention may also be applied to

CC the development of herbicide tolerant plants, plant tissues, plant seeds

CC and plant cells. The polypeptide is useful as herbicide targets in

CC screening assays to identify potential herbicides and inhibitors of

CC GT1209, GT1354 or GT0946 activity. A compound having herbicidal activity

CC is useful for suppressing the growth of a plant. The newly discovered

CC GT1209, GT1354 or GT0946 genes are essential for seedling growth and

CC development. The present sequence is Arabidopsis thaliana GT1209 gene

CC cDNA coding sequence.

XX Sequence 3236 BP; 863 A; 686 C; 680 G; 1007 T; 0 other;

QY Query Match 93.3%; Score 14; DB 24; Length 3236;

Best Local Similarity 100.0%; Pred. No. 2.8e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2295 ACTACTACATGAGC 2308

Search completed: June 26, 2003, 03:11:46

Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(Without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCRTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	4	US-08-918-148-13	Sequence 13, Appl
2	15	100.0	366	2	US-08-958-201-5	Sequence 5, Appl
3	15	100.0	715	3	US-08-545-809A-11	Sequence 11, Appl
4	14	93.3	606	4	US-08-545-809A-22	Sequence 22, Appl
5	14	93.3	2073	4	US-09-321-017B-404	Sequence 404, Appl
6	13.4	89.3	372	3	US-08-767-128-37	Sequence 37, Appl
7	13.4	89.3	2004	1	US-08-471-033-18	Sequence 18, Appl
8	13.4	89.3	2004	2	US-08-471-044-18	Sequence 18, Appl
9	13.4	89.3	2004	2	US-08-463-483A-18	Sequence 18, Appl
10	13.4	89.3	2004	2	US-08-471-046A-18	Sequence 18, Appl
11	13.4	89.3	2004	2	US-08-470-566B-18	Sequence 18, Appl
12	13.4	89.3	2004	3	US-08-469-334-18	Sequence 18, Appl
13	13.4	89.3	2004	3	US-09-300-529-18	Sequence 18, Appl
14	13.4	89.3	2576	1	US-08-471-033-35	Sequence 35, Appl
15	13.4	89.3	2576	2	US-08-471-044-35	Sequence 35, Appl
16	13.4	89.3	2576	2	US-08-463-483A-35	Sequence 35, Appl
17	13.4	89.3	2576	2	US-08-471-046A-35	Sequence 35, Appl
18	13.4	89.3	2576	2	US-08-470-566B-35	Sequence 35, Appl
19	13.4	89.3	2576	2	US-08-469-334-35	Sequence 35, Appl
20	13.4	89.3	2576	3	US-09-300-529-35	Sequence 35, Appl
21	13.4	89.3	2655	1	US-08-471-033-17	Sequence 17, Appl
22	13.4	89.3	2655	1	US-08-471-033-26	Sequence 26, Appl
23	13.4	89.3	2655	2	US-08-471-044-17	Sequence 17, Appl
24	13.4	89.3	2655	2	US-08-471-044-26	Sequence 26, Appl
25	13.4	89.3	2655	2	US-08-463-483A-17	Sequence 17, Appl
26	13.4	89.3	2655	2	US-08-463-483A-26	Sequence 26, Appl
27	13.4	89.3	2655	2	US-08-471-046A-17	Sequence 17, Appl

28	13.4	89.3	2655	2	US-08-471-046A-26	Sequence 26, Appl
29	13.4	89.3	2655	2	US-08-470-566B-17	Sequence 17, Appl
30	13.4	89.3	2655	2	US-08-470-566B-26	Sequence 26, Appl
31	13.4	89.3	2655	2	US-08-469-334-17	Sequence 17, Appl
32	13.4	89.3	2655	2	US-08-469-334-26	Sequence 26, Appl
33	13.4	89.3	2655	3	US-09-300-529-17	Sequence 17, Appl
34	13.4	89.3	2655	3	US-09-300-529-26	Sequence 26, Appl
35	13.4	89.3	4031	1	US-08-471-033-49	Sequence 49, Appl
36	13.4	89.3	4031	2	US-08-471-044-49	Sequence 49, Appl
37	13.4	89.3	4031	2	US-08-463-483A-49	Sequence 49, Appl
38	13.4	89.3	4031	2	US-08-470-566B-49	Sequence 49, Appl
39	13.4	89.3	4031	2	US-08-470-566B-49	Sequence 49, Appl
40	13.4	89.3	4031	2	US-08-469-334-49	Sequence 49, Appl
41	13.4	89.3	4031	3	US-09-300-529-49	Sequence 49, Appl
42	13	86.7	464	2	US-08-449-287-9	Sequence 9, Appl
43	13	86.7	464	2	US-08-449-287-11	Sequence 11, Appl
44	13	86.7	777	2	US-08-860-882A-28	Sequence 28, Appl
45	13	86.7	777	4	US-09-011-769A-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-918-148-13
Sequence 13, Application US/08918148A

Patent No. 6342220

GENERAL INFORMATION:
APPLICANT: Adams, Camellia

APPLICANT: W.

APPLICANT: Carter, Paul J.

APPLICANT: Fendly, Brian M.

APPLICANT: Gurney, Austin L.

TITLE OF INVENTION: Agonist Antibodies

FILE REFERENCE: P0979

CURRENT APPLICATION NUMBER: US/08/918,148A

CURRENT FILING DATE: 1997-08-25

NUMBER OF SEQ ID NOS: 79

SEQ ID NO 13

LENGTH: 15

TYPE: DNA

ORGANISM: artificial

FEATURE:
NAME/KEY: SE5scfv VH CDR1

LOCATION: 1-15

OTHER INFORMATION:

US-08-918-148-13

Query Match

Best Local Similarity 100.0%, Score 15; DB 4; Length 15;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15

DB 1 GACTACTACATGAGC 15

RESULT 2

US-08-958-201-5

Sequence 5, Application US/08958201

Patent No. 5977319

GENERAL INFORMATION:

APPLICANT: Pope, Anthony R

APPLICANT: Pritchard, Kevin

APPLICANT: Williams, Andrew J

TITLE OF INVENTION: Specific binding members for estradiol;

TITLE OF INVENTION: materials and methods

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESSES:

ADDRESS: Marshall O'Toole Gerstein Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

```
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/958,201
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,897
FILING DATE: 21-OCT-1996
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 2G
FEATURE:
NAME/KEY: CDS
LOCATION: 1..366
US-08-958-201-5

Query Match          100.0%; Score 15; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy      1 GACTACTACATGAGC 15
        |||
Db      91 GACTACTACATGAGC 105

RESULT 3
US-08-545-809A-11
; Sequence 11, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; NUMBER OF SEQUENCES: 145
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-11

Query Match          100.0%; Score 15; DB 3; Length 715;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy      1 GACTACTACATGAGC 15
        |||
Db      450 GACTACTACATGAGC 464

RESULT 4
US-08-545-809A-22
; Sequence 22, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; NUMBER OF SEQUENCES: 145
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: human lymphoblast
; CELL LINE: CGM1
US-08-545-809A-22

Query Match          93.3%; Score 14; DB 3; Length 606;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

OY 2 ACTACTACATGAGC 15
|||||
Db 336 ACTACTACATGAGC 349

RESULT 5
US-09-221-017B-404
; Sequence 404, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 404:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2073 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORGANISM: PORPHYROMONAS GINGIVALIS
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...2073
; US-09-221-017B-404

Query Match 93.3%; Score 14; DB 4; Length 2073;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GACTACTACATGAG 14
Db 1828 GACTACTACATGAG 1841

RESULT 6
US-08-767-128-37
; Sequence 37, Application US/08767128
; Patent No. 6111079
; GENERAL INFORMATION:
; APPLICANT: WYLIE, DWANE E.
; APPLICANT: LOPEZ, OSVALDO
; APPLICANT: MURRAY, PETER JOSEPH
; APPLICANT: GOEBEL, PETER
; TITLE OF INVENTION: LEAD BINDING POLYPEPTIDES AND
; TITLE OF INVENTION: NUCLEOTIDES CODING THEREFORE
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6111079west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/767,128
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE: 04-DEC-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09258
; FILING DATE: 05-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/541,373
; FILING DATE: 10-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/462,798
; FILING DATE: 05-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G.
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.49USP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/371-5278
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...372
; OTHER INFORMATION:
; US-08-767-128-37

Query Match 89.3%; Score 13.4; DB 3; Length 372;
Best Local Similarity 93.3%; Pred. No. 69;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GACTACTACATGAG 15

Db 91 GATTACTACATGAGC 105

RESULT 7
US-08-471-033-18

; Sequence 18, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note="Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-033-18

Query Match 89.3%; Score 13.4; DB 1; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15

Db 1381 GACTACTACATGAGC 1395

RESULT 8
US-08-471-044-18

; Sequence 18, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note="Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-044-18

Query Match 89.3%; Score 13.4; DB 2; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;

Matches 14: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GACTACTACATGAGC 15
|||||
Db 1381 GACTACTACATGAGC 1395

RESULT 9
US-08-463-483A-18
Sequence 18, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CTBA-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spull, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
US-08-463-483A-18
Query Match 89.3%; Score 13.4; DB 2; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;

Matches 14: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GACTACTACATGAGC 15
|||||
Db 1381 GACTACTACATGAGC 1395
RESULT 10
US-08-471-046A-18
Sequence 18, Application US/08471046A
Patent No. 586326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 586326artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLv4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8687
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"

US-08-471-046A-18

Query Match 89.3%; Score 13.4; DB 2; Length 2004;

Best Local Similarity 93.3%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 1381 GACTACTACATGAGC 1395

RESULT 11

US-08-470-566B-18

Sequence 18, Application US/08470566B

Patent No. 5872212

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W

APPLICANT: Koziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalini M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No. 5872212el Pesticidal proteins and strains

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5872212artls Corporation

STREET: 3054 Cornwallis Road

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/470,566B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/463,483

FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4

TELEPHONE: 919-541-8689

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2004 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..2004

OTHER INFORMATION: /note="Maize optimized DNA

OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"

US-08-470-566B-18

Query Match 89.3%; Score 13.4; DB 2; Length 2004;

Best Local Similarity 93.3%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 1381 GACTACTACATGAGC 1395

RESULT 12

US-08-469-334-18

Sequence 18, Application US/08469334

Patent No. 5990383

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W

APPLICANT: Koziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalini M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No. 5990383el Pesticidal proteins and strains

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,334

FILING DATE: 06-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/463,483

FILING DATE:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Spivill, W. Murray

REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: CGC 1695/CIP3

TELEPHONE: 919-541-8615

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2004 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-469-334-18

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 2; Length 2004;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACTACTACATGAC 15
DB 1381 GACTACTACATGAC 1395

RESULT 13
US-09-300-529-18
Sequence 18, Application US/09300529
Patent No. 6066783
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziele, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6066783artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/300,529
FILING DATE: TBA
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,334
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19506L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
```

```
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-09-300-529-18

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 3; Length 2004;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACTACTACATGAC 15
DB 1381 GACTACTACATGAC 1395

RESULT 14
US-08-471-033-35
Sequence 35, Application US/08471033
Patent No. 5770696
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziele, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5770696el pesticidal proteins and strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NJ
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..2564
OTHER INFORMATION: /note= "Maize optimized sequence
encoding VIP1(a) with the Bacillus secretion signal removed
US-08-471-033-35
OTHER INFORMATION: contained in pcIB5526"

Query Match 89.3%; Score 13.4; DB 1; Length 2576;
Best Local Similarity 93.3%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
DB 1944 GACTACTACATGAGC 1958

RESULT 15
US-08-471-044-35
Sequence 35, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CI1P3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582

TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..2564
OTHER INFORMATION: /note= "Maize optimized sequence
encoding VIP1(a) with the Bacillus secretion signal removed
US-08-471-044-35
OTHER INFORMATION: contained in pcIB5526"

Query Match 89.3%; Score 13.4; DB 2; Length 2576;
Best Local Similarity 93.3%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
DB 1944 GACTACTACATGAGC 1958

Search completed: June 26, 2003, 04:52:15
Job time : 24.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(Without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published.Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCN_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	393	9	US-09-988-115A-52
2	14	93.3	1422	10	US-09-917-800A-1490
3	14	93.3	2000	9	US-09-938-842A-3563
4	14	93.3	2751	10	US-09-935-943-3
5	14	93.3	3236	10	US-09-935-943-25
6	14	93.3	8002	10	US-09-935-943-10
7	13.4	89.3	211	10	US-09-770-696-95
8	13.4	89.3	470	9	US-09-864-761-175
9	13.4	89.3	467	9	US-09-918-995-9782
10	13.4	89.3	1571	9	US-10-029-180-63
11	13.4	89.3	1571	10	US-09-925-301-93
12	13.4	89.3	465337	10	US-09-933-267A-1
13	13.4	89.3	3309400	9	US-09-738-626-1
14	13	86.7	144	10	US-09-864-761-23136
15	13	86.7	482	10	US-09-864-761-6421
16	13	86.7	1356	10	US-09-815-242-6624
17	13	86.7	1959	9	US-10-102-524-1862
18	13	86.7	3181	9	US-10-098-841-282
19	13	86.7	3743	9	US-10-102-524-1769

20	13	86.7	5248	9	US-09-822-846-76	Sequence 76, Appl
21	13	86.7	5263	9	US-09-374-046A-19	Sequence 19, Appl
22	13	86.7	5263	9	US-10-102-524-1847	Sequence 1847, Ap
23	13	86.7	11597	10	US-09-070-927A-222	Sequence 222, Ap
24	12.6	84.0	1531	9	US-10-106-698-2012	Sequence 2012, Ap
25	12.4	82.7	25	9	US-10-098-263B-1584	Sequence 1584, Ap
26	12.4	82.7	91	10	US-09-969-373-152	Sequence 152, Ap
27	12.4	82.7	179	9	US-09-754-853A-281	Sequence 281, Ap
28	12.4	82.7	203	10	US-09-878-574-13030	Sequence 13030, A
29	12.4	82.7	220	10	US-09-998-598-2246	Sequence 2246, Ap
30	12.4	82.7	263	9	US-09-954-531-657	Sequence 657, Ap
31	12.4	82.7	263	10	US-09-969-347-238	Sequence 238, Ap
32	12.4	82.7	265	10	US-09-878-574-6411	Sequence 6411, Ap
33	12.4	82.7	312	9	US-10-066-543-1572	Sequence 1572, Ap
34	12.4	82.7	393	9	US-09-918-995-35180	Sequence 35180, A
35	12.4	82.7	410	9	US-09-918-995-34814	Sequence 34814, A
36	12.4	82.7	429	7	US-08-781-986A-1087	Sequence 1087, Ap
37	12.4	82.7	442	10	US-09-905-243-42	Sequence 42, Appl
38	12.4	82.7	446	9	US-09-796-692-7562	Sequence 7562, Ap
39	12.4	82.7	446	9	US-10-040-862-7562	Sequence 7562, Ap
40	12.4	82.7	469	10	US-09-864-761-15054	Sequence 15054, A
41	12.4	82.7	494	10	US-09-864-761-3642	Sequence 3642, Ap
42	12.4	82.7	530	10	US-09-879-536-682	Sequence 682, Ap
43	12.4	82.7	587	10	US-09-903-436-16	Sequence 16, Appl
44	12.4	82.7	728	10	US-09-910-943-694	Sequence 694, Ap
45	12.4	82.7	858	10	US-09-764-877-850	Sequence 850, Ap

ALIGNMENTS

RESULT 1
US-09-988-115A-52
Sequence 52, Application US/09988115A
Publication NO. US20030037347A1
GENERAL INFORMATION:
APPLICANT: Robt, James M.
APPLICANT: Goldsby, Richard A.
APPLICANT: Ferguson, Stacy E.
APPLICANT: Kuroiwa, Yoshima
APPLICANT: Tomizuka, Kazuma
APPLICANT: Ishida, Isao
TITLE OF INVENTION: Expression of Xenogenous (human)
FILE REFERENCE: 50195/008003
CURRENT APPLICATION NUMBER: US/09/988,115A
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: US 60/311,625
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 60/256,458
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: US 09/714,185
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: US 60/166,410
PRIOR FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 68
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 52
LENGTH: 393
TYPE: DNA
ORGANISM: Bovine
US-09-988-115A-52

Query Match 100.0%; Score 15; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

```
US-09-917-800A-1490
; Sequence 1490, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1490
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 X13016
US-09-917-800A-1490

Query Match          93.3%; Score 14; DB 10; Length 1422;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAG 14
        |||
Db      347 GACTACTACATGAG 360

RESULT 3
US-09-938-842A-3563/c
; Sequence 3563, Application US/0993842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: S000-3
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3563
```

```
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3563

Query Match          93.3%; Score 14; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ACTACTACATGAGC 15
        |||
Db      1589 ACTACTACATGAGC 1576

RESULT 4
US-09-935-943-3
; Sequence 3, Application US/09935943
; Patent No. US20020120963A1
; GENERAL INFORMATION:
; APPLICANT: Weirich, Lynette M.
; APPLICANT: Budziszewski, Gregory J.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-31509A
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: PCT/EP01/08910
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US 60/222,779
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2751
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2751)
US-09-935-943-3

Query Match          93.3%; Score 14; DB 10; Length 2751;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ACTACTACATGAGC 15
        |||
Db      1998 ACTACTACATGAGC 2011

RESULT 5
US-09-935-943-25
; Sequence 25, Application US/09935943
; Patent No. US20020120963A1
; GENERAL INFORMATION:
; APPLICANT: Weirich, Lynette M.
; APPLICANT: Budziszewski, Gregory J.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-31509A
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: PCT/EP01/08910
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US 60/222,779
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 3236
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (1)..(1056)
US-09-935-943-25

Query Match
Best Local Similarity 93.3%; Score 14; DB 10; Length 3236;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
Db 2295 ACTACTACATGAGC 2308

RESULT 6
US-09-935-943-10
; Sequence 10, Application US/09935943
; Patent No. US20020120963A1
; GENERAL INFORMATION:
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Wegrich, Lynette M.
; APPLICANT: Budziszewski, Gregory J.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-31509A
; CURRENT APPLICATION NUMBER: US/09/935,943
; PRIOR FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: PCT/EP01/08910
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US 60/222,779
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 8002
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-935-943-10

Query Match
Best Local Similarity 93.3%; Score 14; DB 10; Length 8002;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
Db 6239 ACTACTACATGAGC 6252

RESULT 7
US-09-770-696-95/c
; Sequence 95, Application US/09770696
; Patent No. US2001004940A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Ramewka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kriker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2031US (PANA-020PRV)
```

```
; CURRENT APPLICATION NUMBER: US/09/770,696
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,278
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 911
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 211
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-696-95

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 10; Length 211;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
Db 148 GACTACTACATGAGC 134

RESULT 8
US-09-864-761-175
; Sequence 175, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 4917
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
```

```
; SEQ ID NO 175
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034402.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 36
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5e+02
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 34
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2e+02
; OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 1.3e+02
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 46
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 68
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 77
US-09-864-761-175
```

```
Query Match          89.3%; Score 13.4; DB 10; Length 467;
Best Local Similarity 93.3%; Pred. No. 3.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GACTACTACATGAGC 15
Db 184 GACTCCTACATGAGC 198
```

```
RESULT 9
US-09-918-995-9782
; Sequence 9782, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hysed, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-736
; CURRENT APPLICATION NUMBER: US/09/918, 995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9782
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(470)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-9782
```

```
Query Match          89.3%; Score 13.4; DB 9; Length 470;
Best Local Similarity 93.3%; Pred. No. 3.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GACTACTACATGAGC 15
Db 104 GACTACATGAGC 118
```

```
RESULT 10
US-10-029-180-63
; Sequence 63, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffrey C.
; APPLICANT: Trueheart, Josh
```

```
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: No. US20020182708A1 Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029, 180
; PRIOR FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257, 431
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-63
```

```
Query Match          89.3%; Score 13.4; DB 9; Length 1197;
Best Local Similarity 93.3%; Pred. No. 3.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GACTACTACATGAGC 15
Db 754 GACTACTACATGAGC 768
```

```
RESULT 11
US-09-925-301-93/C
; Sequence 93, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925, 301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 1571
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1497)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1516)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1530)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1546)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1571)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-93
```

```
Query Match          89.3%; Score 13.4; DB 10; Length 1571;
Best Local Similarity 93.3%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GACTACTACATGAGC 15
Db 899 GACTACTACATGAGC 885
```

```

RESULT 12
US-09-933-267A-1
, Sequence 1, Application US/09933267A
, Patent No. US20020123095A1
, GENERAL INFORMATION:
, APPLICANT: Kalush, Francis et al.
, TITLE OF INVENTION: Estrogen receptor alpha variants and
, TITLE OF INVENTION: methods of detection thereof
, FILE REFERENCE: CL000258C14
, CURRENT APPLICATION NUMBER: US/09/933,267A
, CURRENT FILING DATE: 2001-08-21
, PRIOR APPLICATION NUMBER: 60/160626
, PRIOR FILING DATE: 1999-10-20
, PRIOR APPLICATION NUMBER: 60/183756
, PRIOR FILING DATE: 2000-02-22
, PRIOR APPLICATION NUMBER: 09/692414
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: 09/768184
, PRIOR FILING DATE: 2001-01-24
, PRIOR APPLICATION NUMBER: 09/804076
, PRIOR FILING DATE: 2001-03-13
, PRIOR APPLICATION NUMBER: 09/826314
, PRIOR FILING DATE: 2001-04-05
, NUMBER OF SEQ ID NOS: 2
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 1
, LENGTH: 465237
, TYPE: DNA
, ORGANISM: human
, US-09-933-267A-1

```

```
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P51957, EVALUO 4.00e-23
; OTHER INFORMATION: EST_HUMAN HIT: A0125616.1, EVALUO 3.00e-73
; OTHER INFORMATION: NT HIT: g111435215, EVALUO 8.00e-76
US-09-864-761-23136

Query Match      86.7%; Score 13; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 CTACTACATGAGC 15
DB      96 CTACTACATGAGC 108

RESULT 15
US-09-864-761-6421
; Sequence 6421, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6421
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006254.10
```

```
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
US-09-864-761-6421

Query Match      86.7%; Score 13; DB 10; Length 482;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 CTACTACATGAGC 15
DB      420 CTACTACATGAGC 432

Search completed: June 26, 2003, 04:56:59
Job time : 68.2 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgac 15

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST :
1: em_estba : *
2: em_esthum : *
3: em_estlin : *
4: em_estnu : *
5: em_estov : *
6: em_estpl : *
7: em_estro : *
8: em_hlc : *
9: gb_est1 : *
10: gb_est2 : *
11: gb_hlc : *
12: gb_est3 : *
13: gb_est4 : *
14: gb_est5 : *
15: em_estfun : *
16: em_estom : *
17: gb_gss : *
18: em_gss_hum : *
19: em_gss_inv : *
20: em_gss_pln : *
21: em_gss_vit : *
22: em_gss_fun : *
23: em_gss_mam : *
24: em_gss_mus : *
25: em_gss_other : *
26: em_gss_pro : *
27: em_gss_rtd : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	308	10	AM801084 MR3-UM006
2	15	100.0	487	10	AM403862 UI-HF-BK0
3	15	100.0	669	12	BG686716 602650678
4	15	100.0	942	14	B0710859 AGENCOURT
5	15	100.0	1112	12	BG760090 602733346
6	15	100.0	1164	12	BF974771 602245420

C	7	14	93.3	230	9	AV272811	AV272811
C	8	14	93.3	250	9	AJ237250	AJ237250
C	9	14	93.3	251	9	AV272672	AV272672
C	10	14	93.3	360	14	C50628	C50628
C	11	14	93.3	377	14	C50318	C50318
C	12	14	93.3	405	12	BF919229	BF919229
C	13	14	93.3	410	17	AO019173	AO019173
C	14	14	93.3	417	12	BF816356	BF816356
C	15	14	93.3	420	9	AA702148	AA702148
C	16	14	93.3	442	17	BH522357	BH522357
C	17	14	93.3	451	17	BH565724	BH565724
C	18	14	93.3	455	9	AL371592	AL371592
C	19	14	93.3	463	9	AL371592	AL371592
C	20	14	93.3	468	17	AO133642	AO133642
C	21	14	93.3	493	13	BM525120	BM525120
C	22	14	93.3	499	13	BM525120	BM525120
C	23	14	93.3	512	13	BM524599	BM524599
C	24	14	93.3	514	17	AO396842	AO396842
C	25	14	93.3	518	12	BG664310	BG664310
C	26	14	93.3	526	9	AL370601	AL370601
C	27	14	93.3	537	12	BE289546	BE289546
C	28	14	93.3	543	12	BG664580	BG664580
C	29	14	93.3	554	13	BG956385	BG956385
C	30	14	93.3	556	12	BG665152	BG665152
C	31	14	93.3	590	9	A1855171	A1855171
C	32	14	93.3	595	10	BE425840	BE425840
C	33	14	93.3	631	12	BG669077	BG669077
C	34	14	93.3	637	10	BE425828	BE425828
C	35	14	93.3	649	13	BJ234400	BJ234400
C	36	14	93.3	661	11	AY067288	AY067288
C	37	14	93.3	661	14	B0507135	B0507135
C	38	14	93.3	677	12	BG597124	BG597124
C	39	14	93.3	688	13	BJ305740	BJ305740
C	40	14	93.3	742	10	BE383112	BE383112
C	41	14	93.3	746	14	B0804288	B0804288
C	42	14	93.3	773	14	B0804369	B0804369
C	43	14	93.3	814	17	AG050744	AG050744
C	44	14	93.3	835	14	B0507136	B0507136
C	45	14	93.3	857	12	BF693325	BF693325

ALIGNMENTS

RESULT 1
LOCUS AM801084 308 bp mRNA linear EST 16-MAY-2000
DEFINITION MR3-UM0065-030300-107-a09 UM0065 Homo sapiens CDNA, mRNA sequence.
ACCESSION AM801084
VERSION AM801084.1 GI:7852954
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 308)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?fl=st2=MR3-DM0065-030
 (http://www.ludwig.org.br/scripts/gethtml2.pl?fl=st2=MR3-DM0065-030
 Seq primer: puc 18 forward
 High quality sequence stop: 308.
 Location/Qualifiers
 1. 308

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="DM0065"
 /dev_stage="Adult"
 /note="Organ: uterus; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT
 65 a 89 c 81 g 73 t
 ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 308;
 Best Local Similarity 100.0%; Pred. No. 9 8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GACTACTACATGAGC 15
 ||||||||||||||||
 Db 190 GACTACTACATGAGC 176

RESULT 2
 LOCUS AM403862 487 bp mRNA linear EST 16-FEB-2000
 DEFINITION UT-HF-BK0-abc-a-12-0-UT.r1 NIH_MGC_36 Homo sapiens cDNA clone
 IMAGE:3056926 5', mRNA sequence.
 ACCESSION AM403862
 VERSION AM403862.1 GI:6923015
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 487)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab
 cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward.
 Location/Qualifiers
 1. 487

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3056926"
 /clone_lib="NIH_MGC_36"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /note="Vector: pRT3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima

BASE COUNT
 109 a 115 c 143 g 120 t
 ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 487;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GACTACTACATGAGC 15
 ||||||||||||||||
 Db 215 GACTACTACATGAGC 229

RESULT 3
 LOCUS BG686716 669 bp mRNA linear EST 01-MAY-2001
 DEFINITION 602650678F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4763274 5',
 IMAGE:4763274 5',
 ACCESSION BG686716
 VERSION BG686716.1 GI:13918113
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 669)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov plate: LCM1618 row: k column: 19
 High quality sequence stop: 665.
 Location/Qualifiers
 1. 669

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4763274"
 /clone_lib="NIH_MGC_48"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
 BASE COUNT
 149 a 184 c 186 g 150 t
 ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 669;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GACTACTACATGAGC 15
 ||||||||||||||||
 Db 217 GACTACTACATGAGC 231

RESULT 4
 LOCUS BG710859 942 bp mRNA linear EST 16-JUL-2002
 DEFINITION AGENCOURT_8501500 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6303061

5', mRNA sequence.
 B0710859
 VERSION B0710859.1 GI:21849758
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 942)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Mark Watson
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2521 row: 1 column: 14
 High quality sequence stop: 510.

FEATURES
 source Location/Qualifiers
 1..942
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6303061"
 /lab_host="NIH_MGC_113"
 /note="Organ: spleen; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT
 197 a 291 c 263 g 191 t

ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 942;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
 |||
 |||

Db 208 GACTACTACATGAGC 222

RESULT 5
 B0760090 1112 bp mRNA linear EST 15-MAY-2001
 LOCUS 60273346F1 NIH_MGC_48 Homo sapiens CDNA clone IMAGE:4878851 5',
 DEFINITION mRNA sequence.
 B0760090
 ACCESSION B0760090.1 GI:14070743
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1112)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1765 row: k column: 12
 High quality sequence stop: 509.
 Location/Qualifiers
 1..1112
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4878851"
 /clone_lib="NIH_MGC_48"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT
 346 a 268 c 318 g 180 t

ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 1112;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
 |||
 |||

Db 218 GACTACTACATGAGC 232

RESULT 6
 BF974771 1164 bp mRNA linear EST 22-JAN-2001
 LOCUS 602245420F1 NIH_MGC_48 Homo sapiens CDNA clone IMAGE:4336541 5',
 DEFINITION mRNA sequence.
 BF974771
 ACCESSION BF974771.1 GI:12341986
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1164)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1209 row: g column: 06
 High quality sequence stop: 696.
 Location/Qualifiers
 1..1164
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4336541"
 /clone_lib="NIH_MGC_48"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling

KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida; Rhabditioidea
; Rhabditidae; Peloderinae: Caenorhabditis.
AUTHORS Kohara, Y., Motoshashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yui Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
location/Qualifiers

FEATURES
SOURCE 1. 377
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk480c3"
/clone_lib="Yui Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev-stage="embryo"
BASE COUNT 112 a 76 c 86 g 102 t 1 others
ORIGIN

Query Match 93.3%; Score 14; DB 14; Length 377;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ACTACTACATGAGC 15
|||||
Db 114 ACTACTACATGAGC 127

RESULT 12 405 bp mRNA linear EST 19-JAN-2001
BF919229
LOCUS CMO-NT0131-281000-641-g10 NT0131 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF919229
ACCESSION BF919229.1 GI:12315117
VERSION EST.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 405)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CMO<2=CMO-NT0131-
281000-641-g10<3=2000-10-28<4=1)

Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 405.
Location/Qualifiers

FEATURES
SOURCE 1. 405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0131"
/dev-stage="Adult"
/note="Organ: nervous tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI: A mini-library was made by cloning products
derived from ORESSES PCR (U.S. letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 125 a 88 c 57 g 135 t
ORIGIN

Query Match 93.3%; Score 14; DB 12; Length 405;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ACTACTACATGAGC 15
|||||
Db 258 ACTACTACATGAGC 271

RESULT 13 410 bp DNA linear GSS 09-JUN-1998
AQ019173
LOCUS CIT-HSP-2303C19.TR CIT-HSP Homo sapiens genomic clone 2303C19, DNA
DEFINITION sequence.
ACCESSION AQ019173
VERSION AQ019173.1 GI:3197909
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 410)
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., White, C., Shizuya, H.,
Simon, M. and Venter, J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2303C19.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hunguen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers

FEATURES
SOURCE 1. 410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2303C19"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 123 a 91 c 56 g 140 t
ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-25

Perfect score: 15

Sequence: 1 acctacgcatgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenBank: *
2: gb_da: *
3: gb_htg: *
4: gb_in: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rnd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	6	AR183486 Sequence
2	15	100.0	370	9	AF029333 Homo sapi
3	15	100.0	717	9	AF048775 Homo sapi
4	15	100.0	2030	9	AK090859 Homo sapi
5	15	100.0	42845	9	HS073633 Human chrom
6	15	100.0	114287	9	AP000590 Homo sapi
7	15	100.0	157226	9	AC018902 Homo sapi
8	15	100.0	194201	2	AP000896 Homo sapi
9	15	100.0	210402	9	AP001271 Homo sapi
10	15	100.0	214415	9	AC010674 Homo sapi
11	15	100.0	333800	1	SME591792 Sinorhizo
12	14	93.3	787	9	HS1GVDJ2
13	14	93.3	38891	2	AC010298 Homo sapi
14	14	93.3	42432	1	AF083072 Homo sapi
15	14	93.3	72354	9	AL357123 Human DNA
16	14	93.3	83629	2	AC110485 Rattus no
17	14	93.3	110157	2	AC130966 Rattus no
18	14	93.3	151438	2	AC103516 Rattus no
19	14	93.3	152835	8	AP004224 Oryza sat
20	14	93.3	154681	2	AC022000 Homo sapi
21	14	93.3	159677	2	AC113802 Rattus no
22	14	93.3	161722	2	AC121969 Mus muscu
23	14	93.3	162742	2	AC096676 Homo sapi
24	14	93.3	164764	2	AC102957 Rattus no
25	14	93.3	176542	2	AC097249 Rattus no
26	14	93.3	221380	2	AC094347 Rattus no
27	14	93.3	341720	2	AC126939 Mus muscu
28	13.4	89.3	231	9	HS407974 Homo sapi
29	13.4	89.3	242	9	HS402495 Homo sapi
30	13.4	89.3	273	9	HS408229 Homo sapi
31	13.4	89.3	288	9	HS280501 H. sapiens B
32	13.4	89.3	290	9	AF077438 Homo sapi
33	13.4	89.3	305	9	HS280655 H. sapiens B
34	13.4	89.3	308	9	AY003764 Homo sapi
35	13.4	89.3	311	9	AY003712 Homo sapi
36	13.4	89.3	329	9	AY003723 Homo sapi
37	13.4	89.3	329	9	AY003765 Homo sapi
38	13.4	89.3	346	9	HS1GVBAB Homo sapien
39	13.4	89.3	354	9	HS4132850 Homo sapi
40	13.4	89.3	354	9	HS4271549 Homo sapi
41	13.4	89.3	357	9	AJ132851 Homo sapi
42	13.4	89.3	363	9	AJ245298 Homo sapi
43	13.4	89.3	363	9	HS4245298 Homo sapi
44	13.4	89.3	363	9	HS4245300 Homo sapi
45	13.4	89.3	372	6	AR160944 Sequence

ALIGNMENTS

RESULT 1
LOCUS AR183486 15 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 25 from patent US 6342220.
ACCESSION AR183486
VERSION AR183486.1 GI:20227455
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 25 29-JAN-2002;
FEATURES Location/Qualifiers

source	1. .15	/organism="unknown"										
BASE COUNT	4 a	6 c	3 g	2 t								
ORIGIN												
Query Match	100.0%;	Score 15;	DB 6;	Length 15;								
Best Local Similarity	100.0%;	Pred. No. 2.5e+04;										
Matches	15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;							
QY	1 ACCTACGGCATTGCAC	15										
Db	1 ACCTACGGCATTGCAC	15										
RESULT 2												
AF029333	370 bp	mRNA	linear	PRI 10-NOV-1997								
LOCUS												
DEFINITION	Homo sapiens puu1 immunoglobulin heavy chain V region	mRNA,										
ACCESSION	AF029333	partial cds.										
VERSION	AF029333.1	GI:2599529										
KEYWORDS												
ORGANISM	Homo sapiens.											
SOURCE	Homo sapiens											
REFERENCE	EuroArytola, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 370)											
AUTHORS	Salonen, E.-M., Parren, P.W.H.I., Graus, Y.F., Lundkvist, A., Fisticaro, P., Vapalahti, O., Kallio-Kokko, H., Vaheri, A. and Burton, D.R.											
TITLE	Human recombinant Puumala virus antibodies: cross reaction with other hantaviruses and use in diagnostics											
JOURNAL	Unpublished											
REFERENCE	2 (bases 1 to 370)											
AUTHORS	Salonen, E.-M., Parren, P.W.H.I., Graus, Y.F., Lundkvist, A., Fisticaro, P., Vapalahti, O., Kallio-Kokko, H., Vaheri, A. and Burton, D.R.											
TITLE	Direct Submission											
JOURNAL	Submitted (09-OCT-1997) Immunology, The Scripps research Institute											
FEATURES	10550 N. Torrey Pines Road, La Jolla, CA 92037, USA											
source	Location/Qualifiers											
CDS	1. .370	/organism="Homo sapiens"										
	/isolate="puu1"											
	/db_xref="taxon:9606"											
	<1. .>370											
	/note="Pd fragment: recognizes Puumala virus G2 glycoprotein (G2c epitope)"											
	/product="immunoglobulin heavy chain V region"											
	/codon_start=1											
	/protein_id="AAB84156.1"											
	/db_xref="GI:2599530"											
	/translation="TLESGGVVOPGRSRRLSCVAGFTPTTGGHMYROADGKLEFVATISVOSGOKDVAISLKGRTFISDNKSTIVYLQMSLRLRADIATVYICAAQENFYD											
	RGATIHMGSGSLVIYSSASTK"											
BASE COUNT	84 a	102 c	103 g	81 t								
ORIGIN												
Query Match	100.0%;	Score 15;	DB 9;	Length 370;								
Best Local Similarity	100.0%;	Pred. No. 8.5e+03;										
Matches	15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;							
QY	1 ACCTACGGCATTGCAC	15										
Db	82 ACCTACGGCATTGCAC	96										
RESULT 3												
AF048775	717 bp	mRNA	linear	PRI 31-DEC-1996								
LOCUS												
DEFINITION	Homo sapiens clone 12B5 anti-mp1 scfv mRNA, partial cds.											
ACCESSION	AF048775											
VERSION	AF048775.1	GI:2911501										

KEYWORDS	Homo sapiens.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 717)
AUTHORS	Merchant,A.M., Zhu,Z., Yuan,J.Q., Goddard,A., Adams,C.W., Presta,L.G. and Carter,P.
TITLE	An efficient route to human bispecific IgG
JOURNAL	Nat. Biotechnol. 16 (7), 677-681 (1998)
MEDLINE	98325681
PUBMED	9661204
REFERENCE	2 (bases 1 to 717)
AUTHORS	Goddard,A., Yuan,J.Q., Zhu,Z. and Carter,P.
TITLE	Direct Submission
JOURNAL	Submitted (17-FEB-1998) Molecular Oncology, Genentech Inc, 1 DNA Way, South San Francisco, CA 94080, USA
FEATURES	Location/Qualifiers
source	1..717
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="12B5"
	<1..>717
CDS	/codon_start=1
	/product="anti-Mp1 scFv"
	/protein_id="AAC98736.1"
	/db_xref="GI:2911502"
	/translation="OVOLVSGGGLVPRGGSLSCAVSGITLTRTYGMHVRQAPRGKGL LEWVYISPDGRSEHYVADSVGGRFTISRDSKNTLYLOMNSLRADYAVVYCARGAHYH GFDIWGGCTWYVSSGGCGSGCGGSDIOMTOSPSRTISACDRIYVTCARSAH YTHLWAMIQCKPCKAPKRLLIYKASLSLGSAPRSFGSGSDITFLTITSLDPPDFATY YCOOYSNPITPFGGTYKLEIK"
BASE COUNT	161 a 188 c 215 g 153 t
ORIGIN	
Query Match	100.0%; Score 15; DB 9; Length 717;
Best Local Similarity	100.0%; Pred. No. 6,8e+03;
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ACTACGGCATGCAC 15
Db	91 ACTACGGCATGCAC 105
RESULT 4	
AK090859/c	2030 bp mRNA linear PRI 15-JUL-2002
DEFINITION	Homo sapiens cDNA FLJ3540 f1s, clone BRAMY2007613.
ACCESSION	AK090859
VERSION	AK090859.1 GI:21745098
KEYWORDS	oligo capping; f1s (full insert sequence).
SOURCE	Homo sapiens amygdala cDNA to mRNA, clone.lib:BRAMY2 clone:BRAMY2007613.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1
AUTHORS	Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,T., Komiyama,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wadatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahata,K., Masuno,Y., Nagai,K. and Isogai,T.
TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2030)
AUTHORS	Isogai,T. and Yamamoto,J.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FlJ Project(HRI Team): 2-6-7 Kazusa Kametani, Kisarazu, Chiba 292-0812, Japan (E-mail:igenomats@r1.co.jp;tel:01-438-52-3975; fax:01-438-52-3986

COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
source
1. .2030
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BRAMY2007613"
/tissue_type="amygdala"
/clone_lib="BRAMY2"
/note="cloning vector: PME18SFL3"

BASE COUNT 586 a 345 c 402 g 697 t

ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 2030;
Best Local Similarity 100.0%; Pred. No. 4; de+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCAGTCAC 15
|||||
361 ACCTACGGCAGTCAC 347

Db

RESULT 5
HSU73633/c 42845 bp DNA linear PRI 19-JUN-1997
LOCUS Human chromosome 11 146h12 cosmid, complete sequence.
DEFINITION U73633
ACCESSION U73633
VERSION U73633.1 GI:2199506
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 42845)
Evans,G.A., Bradbury,P., Brignac,S., Bumeister R., Davie,J., Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hanher,L., Harris,J., Hinson,S., McFarland,J., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S., Schegeman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
Direct Submission
Submitted (07-OCT-1996) McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas 5333 Harry Hines Blvd, Dallas, TX 75235-8591, USA
On Jun 19, 1997 this sequence version replaced gi:1737188.
Location/Qualifiers
1. .42845
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23"
/clone="146h12"
/clone_lib="library prepared from flow sorted human chromosome 11 derived from Chinese hamster ovary (CHO) monochromosomal somatic cell hybrid, J1; vector: scos-1"
2218. .2488
/rpt_family="ALU"
/rpt_type-dispersed
complement(12155. .12233)
/rpt_family="MIR"
/rpt_type-dispersed
12594. .12879
/rpt_family="ALU"
/rpt_type-dispersed
13713. .14193
/rpt_family="ALU"

TITLE
JOURNAL

COMMENT
FEATURES
source
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

repeat_region
14374. .14661
/rpt_family="ALU"
/rpt_type-dispersed
complement(14849. .15071)
/rpt_family="ALU"
/rpt_type-dispersed
16557. .16630
/rpt_family="MIR"
/rpt_type-dispersed
20477. .20764
/rpt_family="ALU"
/rpt_type-dispersed
complement(22420. .22688)
/rpt_family="ALU"
/rpt_type-dispersed
complement(24053. .24356)
/rpt_family="ALU"
/rpt_type-dispersed
26266. .26540
/rpt_family="ALU"
/rpt_type-dispersed
complement(27183. .27512)
/rpt_family="MLT1"
/rpt_type-dispersed
complement(28625. .28911)
/rpt_family="ALU"
/rpt_type-dispersed
complement(32586. .32629)
/rpt_family="L1"
/rpt_type-dispersed
complement(32743. .33041)
/rpt_family="ALU"
/rpt_type-dispersed
complement(33028. .33172)
/rpt_family="L1"
/rpt_type-dispersed
35218. .35473
/rpt_family="ALU"
/rpt_type-dispersed
complement(36781. .37041)
/rpt_family="MLT1"
/rpt_type-dispersed
39388. .39646
/rpt_family="ALU"
/rpt_type-dispersed
complement(39724. .39766)
/rpt_family="MER46"
/rpt_type-dispersed
complement(39985. .40055)
/rpt_family="MER46"
/rpt_type-dispersed

BASE COUNT 9620 a 11179 c 11246 g 10800 t

ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 42845;
Best Local Similarity 100.0%; Pred. No. 1; de+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCAGTCAC 15
|||||
1464 ACCTACGGCAGTCAC 1450

Db

RESULT 6
AP000590 114287 bp DNA linear PRI 18-JUL-2001
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone: CMB9-52H4, complete
DEFINITION
ACCESSION AP000590
VERSION AP000590.6 GI:14861095
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone: CMB9-52H4.

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Homo sapiens genomic DNA
JOURNAL      Published Only in Database (1999)
REFERENCE
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
              1-7-22 Saitoh-Chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gscc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
              On Jul 17, 2001 this sequence version replaced gi:10880434.
FEATURES
SOURCE
Location/Qualifiers
1..114287
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CMB9-52H4"
BASE COUNT    28585 a 30994 c 28832 g 25876 t
ORIGIN
Query Match    100.0%; Score 15; DB 9; Length 114287;
Best Local Similarity 100.0%; Pred. NO. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGCGCATGCAC 15
    |||||
DB 22215 ACCTACGCGCATGCAC 22229

RESULT 7
AC018902      157226 bp  DNA  linear  PRI 09-AUG-2001
LOCUS        Homo sapiens chromosome 15 clone RP11-327C2 map 15q21.2, complete
sequence.
ACCESSION    AC018902
VERSION      AC018902.8  GI:15138554
KEYWORDS     HTG.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S.,
              Bloom,S., Dots,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
              James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
              and Hood,L.
TITLE        Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL      Unpublished
REFERENCE
AUTHORS      2 (bases 1 to 157226)
              Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
              Bloom,S., Dots,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
              James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
              and Hood,L.
TITLE        Direct Submission
JOURNAL      Submitted (22-DEC-1999) Multimegabase Sequencing Center, University
              of Washington, PO BOX 357730, Seattle, WA 98195, USA
              3 (bases 1 to 157226)
              Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
              Burke,D., Dots,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
              Pate,D., and Hood,L.
TITLE        Direct Submission
JOURNAL      Submitted (09-AUG-2001) Multimegabase Sequencing Center, Institute
              for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
              98105, USA
              On Aug 9, 2001 this sequence version replaced gi:13173608.
COMMENT
----- Genome Center

```

```

Center: Multimegabase Sequencing Center
Center code: UMNCS
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leetown@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping BACs AC010674 [drafting center:
UMNCS], AC025917 [drafting center: UMNCS] and AC073438 [drafting
center: WTBRI] were added for finishing
-----
Location/Qualifiers
1..157226
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.2"
/clone="RP11-327C2"
/clone_lib="RP11 human BAC library 11"
/note="Data from overlapping BACs RP11-430B1, RP11-23N2,
and RP11-472P3 were added and the consensus sequence was
determined from RP11-327C2 to the extent possible"
1..37195
/note="overlap with RP11-430B1 AC010674"
80042..157226
/note="overlap with RP11-472P3 AC073438"
122504..122615
/note="low quality data"
148580..157226
/note="overlap with RP11-23N2 AC025917"
BASE COUNT    49939 a 31674 c 30429 g 45184 t
ORIGIN
Query Match    100.0%; Score 15; DB 9; Length 157226;
Best Local Similarity 100.0%; Pred. NO. 1.1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGCGCATGCAC 15
    |||||
DB 1953 ACCTACGCGCATGCAC 1967

RESULT 8
AP000896      194201 bp  DNA  linear  HTG 30-MAY-2000
LOCUS        Homo sapiens chromosome 11 clone RP11-677D19 map 11q13, WORKING
DRAFT SEQUENCE, 61 unordered pieces.
ACCESSION    AP000896.2  GI:8119039
VERSION      AP000896.2  GI:8119039
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Homo sapiens DNA, clone:RP11-677D19.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Sequencing of human chromosome 11 RP11-677D19
JOURNAL      Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
              Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
              Japan (E-mail:hattori@gsr.riken.go.jp,
              URL:http://hgp.gscc.riken.go.jp/, Tel:81-42-778-9923,
              Fax:81-42-778-9924)
              On May 31, 2000 this sequence version replaced gi:6997719.
COMMENT
----- Genome Center

```

```
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp

----- Project Information
Center project name: HumDrat11
Center clone name: RP11-677D19

----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator; ER-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170176 bases at least Q40
Consensus quality: 175910 bases at least Q30
Consensus quality: 181676 bases at least Q20
Insert size: 188201; sum-of-contigs
Quality coverage: 4.42x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
61 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.

1 10439 contig of 10439 bp in length
10540 19520 contig of 8981 bp in length
19621 25661 contig of 6041 bp in length
25762 33211 contig of 7450 bp in length
33312 39008 contig of 5697 bp in length
39109 44927 contig of 5819 bp in length
45028 50794 contig of 5767 bp in length
50895 56215 contig of 5321 bp in length
56316 62353 contig of 6238 bp in length
62654 66956 contig of 4303 bp in length
67057 72631 contig of 5575 bp in length
72732 76352 contig of 5621 bp in length
76453 81462 contig of 5010 bp in length
81563 85655 contig of 4093 bp in length
85756 88387 contig of 2633 bp in length
88488 92170 contig of 3683 bp in length
92271 97116 contig of 4846 bp in length
101637 101536 contig of 4320 bp in length
105192 105091 contig of 3455 bp in length
105192 108459 contig of 3268 bp in length
108560 112286 contig of 3727 bp in length
112387 116086 contig of 3700 bp in length
116187 119641 contig of 3465 bp in length
119742 122405 contig of 2664 bp in length
122506 125633 contig of 3128 bp in length
125734 129128 contig of 3395 bp in length
129229 133328 contig of 3100 bp in length
132429 135207 contig of 2779 bp in length
135308 138331 contig of 3024 bp in length
138432 139316 contig of 885 bp in length
139417 141810 contig of 2394 bp in length
141911 143756 contig of 1846 bp in length
143857 145781 contig of 1925 bp in length
145882 148246 contig of 2365 bp in length
148347 150678 contig of 2532 bp in length
150799 152797 contig of 1819 bp in length
152898 154641 contig of 1744 bp in length
154742 156816 contig of 2075 bp in length
156917 158344 contig of 1428 bp in length
158445 160814 contig of 2370 bp in length
160915 162537 contig of 1623 bp in length
162638 164708 contig of 2071 bp in length
164809 166629 contig of 1821 bp in length
166730 168328 contig of 1599 bp in length
168429 170497 contig of 2069 bp in length
170598 172814 contig of 2217 bp in length
172915 174700 contig of 1786 bp in length
174801 175977 contig of 1177 bp in length
```

```
176078 178046 contig of 1969 bp in length
178147 179644 contig of 1498 bp in length
179745 181427 contig of 1683 bp in length
181528 182786 contig of 1259 bp in length
182887 184440 contig of 1554 bp in length
184541 185555 contig of 1015 bp in length
185656 187161 contig of 1506 bp in length
187262 188593 contig of 1332 bp in length
188694 190012 contig of 1319 bp in length
190113 190464 contig of 352 bp in length
190565 191917 contig of 1353 bp in length
192018 193099 contig of 1082 bp in length
193200 194201 contig of 1002 bp in length

Sequence updated (26-May-2000).
NOTE: This is a 'working draft' sequence. It currently
consists of 61 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 10439: contig of 10439 bp in length
10440 10539: gap of 100 bp
10540 19520: contig of 8981 bp in length
19521 19620: gap of 100 bp
19621 25661: contig of 6041 bp in length
25662 25761: gap of 100 bp
25762 33211: contig of 7450 bp in length
33212 33311: gap of 100 bp
33312 39008: contig of 5697 bp in length
39009 39108: gap of 100 bp
39109 44927: contig of 5819 bp in length
44928 45027: gap of 100 bp
45028 50794: contig of 5767 bp in length
50795 50894: gap of 100 bp
50895 56215: contig of 5321 bp in length
56216 56315: gap of 100 bp
56316 62553: contig of 6238 bp in length
62554 62653: gap of 100 bp
62654 66956: contig of 4303 bp in length
66957 67056: gap of 100 bp
67057 72631: contig of 5575 bp in length
72632 72731: gap of 100 bp
72732 76352: contig of 3621 bp in length
76353 76452: gap of 100 bp
76453 81462: contig of 5010 bp in length
81463 81562: gap of 100 bp
81563 85655: contig of 4093 bp in length
85656 85755: gap of 100 bp
85756 88387: contig of 2632 bp in length
88388 88487: gap of 100 bp
88488 92170: contig of 3683 bp in length
92171 92270: gap of 100 bp
92271 97116: contig of 4846 bp in length
97117 97216: gap of 100 bp
97217 101536: contig of 4320 bp in length
101537 101636: gap of 100 bp
101637 105091: contig of 3455 bp in length
105092 105191: gap of 100 bp
105192 108459: contig of 3268 bp in length
108460 108559: gap of 100 bp
108560 112286: contig of 3727 bp in length
112287 112386: gap of 100 bp
112387 116086: contig of 3700 bp in length
116087 116186: gap of 100 bp
116187 119641: contig of 3455 bp in length
119642 119741: gap of 100 bp
119742 122405: contig of 2664 bp in length
122406 122505: gap of 100 bp
122506 125633: contig of 3128 bp in length
125634 125733: gap of 100 bp
125734 129128: contig of 3395 bp in length
```

```

* 129129 129228: gap of 100 bp
* 129229 132328: contig of 3100 bp in length
* 132329 132428: gap of 100 bp
* 132429 135207: contig of 2779 bp in length
* 135208 135307: gap of 100 bp
* 135308 138331: contig of 3024 bp in length
* 138332 138431: gap of 100 bp
* 138432 139316: contig of 885 bp in length
* 139317 139416: gap of 100 bp
* 139417 141810: contig of 2394 bp in length
* 141811 141910: gap of 100 bp
* 141911 143756: contig of 1846 bp in length
* 143757 143856: gap of 100 bp
* 143857 145781: contig of 1925 bp in length
* 145782 145881: gap of 100 bp
* 145882 148246: contig of 2365 bp in length
* 148247 148346: gap of 100 bp
* 148347 150878: contig of 2532 bp in length
* 150879 150978: gap of 100 bp
* 150979 152797: contig of 1819 bp in length
* 152798 152897: gap of 100 bp
* 152898 154641: contig of 1744 bp in length
* 154642 154741: gap of 100 bp
* 154742 156816: contig of 2075 bp in length
* 156817 156916: gap of 100 bp
* 156917 158344: contig of 1428 bp in length
* 158345 158444: gap of 100 bp
* 158445 160814: contig of 2370 bp in length

Query Match 100.0%; Score 15; DB 2: Length 194201;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCTACGCGCATGCAC 15
Db 59787 ACCTACGCGCATGCAC 59801

RESULT 9
LOCUS AP001271
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone: RP11-826F13,
complete sequence.
ACCESSION AP001271.4 GI:15320468
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Homo sapiens DNA, clone: RP11-826F13.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Saito-chou, Tsukuba, Ibaraki, Japan
(E-mail: hattori@gsr.riken.go.jp, URL: http://hpg.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
On Aug 27, 2001 this sequence version replaced gi:11138090.

FEATURES
source 1..210402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-826F13"

BASE COUNT 50660 a 55101 c 55700 g 48941 t

ORIGIN
Query Match 100.0%; Score 15; DB 9: Length 210402;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCTACGCGCATGCAC 15
Db 210307 ACCTACGCGCATGCAC 210321

RESULT 10
LOCUS AC010674
DEFINITION Homo sapiens chromosome 15 clone RP11-430B1 map 15q21.1, complete
sequence.
ACCESSION AC010674
VERSION AC010674.11 GI:15042780
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Rowen, L., Madan, A., Qin, S., Abbas, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Laskey, S.,
Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
AUTHORS Rowen, L., Madan, A., Qin, S., Abbas, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Laskey, S.,
Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 214415)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (31-JUN-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
On Jul 31, 2001 this sequence version replaced gi:13878269.

COMMENT
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbl.washington.edu/msg_wmw
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

Note: Data from overlapping BACs AC023906 [drafting center:
UWMSC], AC018902 [drafting center: UWMSC], and AC018903 [drafting
center: UWMSC] were added for finishing
-----
Location/Qualifiers
1..214415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.1"
/clone="RP11-430B1"
/clone.lib="RP11-430B1"
/note="Data from overlapping BACs CTD-2184D3, RP11-327C2,
and RP11-337B11 were added and the consensus sequence was
determined from RP11-430B1 to the extent possible"
1..11294
/note="Overlap with CTD-2184D3 AC023906"

misc_feature

```

```

unsure      65643      /note="low quality data"
unsure      77163      /note="low quality data"
unsure      92425      /note="low quality data"
unsure      108954      /note="low quality data"
misc-feature 119015      /note="low quality data"
              119015      /note="sequence data generated from subcloned PCR product
              only"
unsure      126825      /note="low quality data"
unsure      146876      /note="low quality data"
unsure      165370      /note="low quality data"
unsure      177221      /note="low quality data"
misc-feature 177221      /note="overlap with RPL1-327C2"
BASE COUNT  61817 a 47930 c 46801 g 57867 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 214415;
Best Local Similarity 100.0%; Pred. No. 9,9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCTACGCATGCAC 15
Db      179173 ACCTACGCATGCAC 179187

RESULT 11
SME591792      333800 bp      DNA      linear      BCT 05-JUL-2002
LOCUS      Sme591792
DEFINITION      Sinorhizobium meliloti 1021 complete chromosome; segment 11/12.
ACCESSION      AL591792 AL591688
VERSION      AL591792.1 GI:15075850
KEYWORDS
SOURCE
ORGANISM      Sinorhizobium meliloti.
              Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
              Rhizobiaceae; Sinorhizobium.
REFERENCE
AUTHORS      1 (bases 1 to 333800)
              Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
              Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
              Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masny,D.,
              Pohl,T., Portetlelle,D., Puhler,A., Punelle,B., Ramsperger,U.,
              Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
              Analysis of the chromosome sequence of the legume symbiont
              Sinorhizobium meliloti strain 1021
              Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
TITLE
JOURNML      21396507
MEDLINE      11481430
PUBMED      2 (bases 1 to 333800)
REFERENCE
AUTHORS      Gouzy,J.
TITLE      Direct Submission
JOURNML      Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
              EU Consortium
COMMENT
              MELILO EU Consortium:
              Laboratoire de Biologie Moléculaire des Relations
              Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
              France, Laboratoire de Génétique et Développement UMR6061-CNRS,
              Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,
              France, GATC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany,
              Universitt Bielefeld, Biologie IV (Genetik) Universittsstr 25,
              D-33615 Bielefeld, Germany, Unité de Biochimie physiologique,
              Université Catholique de Louvain, Place Croix du Sud 2, Bte 20,
              B-1348 Louvain-la-Neuve, Belgium, Unité de Microbiologie, Faculté
              des Sciences Agronomiques de Gembloux, Avenue Marezchal Juin 6,
              B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
              http://sequence.toulouse.inra.fr/meliloti.html.
FEATURES
Source      1..333800

```

```

gene
CDS
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
complement(233..1114)
/gene="SMC04043"
complement(233..1114)
/gene="SMC04043"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED_HYPOTHETICAL_PROTEIN"
/protein_id="CAC47405.1"
/db_xref="GI:15075851"
/db_xref="SPTREMBL:Q92M30"
/transl_table="MGEAWEFEFEVLEPASQIPFVNSPHSGRHYPOEFLDQSRD
PHSTRSEDFHVELEFRSATFLGAPLRAHPRAFLVNPEDLPMEFGALPRA
NLSMRVAGIGTVPRLVAENMEYRGEPVEQALERTETTKYHATLRKLIATYV
EFGKSTLIDCSHMGONVHLRPGSGRRPPTIGDRYGTSAALSRVAELLEQGYAVA
RKRPYAGFTIEHTGRTPLGLALQIEINRGLYDEATLTKKPFALENDLATFIAA
LARHVEDFGAYLPLAAE"
1305..1667
/gene="SMC04044"
1305..1667
/gene="SMC04044"
/function="miscellaneous: not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE_2-COMPONENT_RECEIVER_DOMAIN_PROTEIN"
/protein_id="CAC47406.1"
/db_xref="GI:15075852"
/db_xref="SPTREMBL:Q92M29"
/transl_table="MTAKILLEDNDNRRLPYKALEKAGYKVLSDNGASAYDLRE
EPEFLNDIVMEKDGIETLARARELPDLKVFITGPAAVALNPSKAPADKAVLS
KPFHLRLDVLNVEKMLAA"
1952..2026
/gene="trna-VAL_GAC"
1952..2026
/gene="trna-VAL_GAC"
/product="trna-VAL"
/note="codon recognized: GUC; predicted by tRNAscan-SE"
/evidence=not_experimental
2459..4297
/gene="11VD2 OR SMC04045"
2459..4297
/gene="11VD2 OR SMC04045"
/EC_number="4.2.1.9"
/function="small molecule metabolism: amino acid
biosynthesis: isoleucine/valine"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE_DIHYDROXY-ACID_DEHYDRATASE_PROTEIN"
/protein_id="CAC47407.1"
/db_xref="GI:15075853"
/db_xref="SPTREMBL:Q92M28"
/transl_table="MPAYRSRTTTHGRNMGAGLWRATGMKDSDFGKPIAVVNSFT

```

```

OFEVGHVHDKDLSGLVARETEPAGGVAKEENTIAVDGDIAMGHDMGLSIPRELIAD
SVEVYNAHCHADAVVCSNCKITPGHIMALRINIPAVYPSGSPMAGVYVHGKTH
ALDLDVAVNAADKRVSDVDQITERSACPCCSGFTFANSNCLTEALGSLPEN
GSTLATIHADRKLFEVAGHLIVDLARRRYEDEDRLVPRNATKQATEENMALDIAG
GSTNVIHLIAAEVGEIDFTMDIDIRLSRVPCLSVAPAKADVHEDEHARAGIIS
ILGELDKGGLINRDCPTVHAETLGDADIRMDITRTSSDYKFRAPAGGIPPOVARS
OEARWDELDTRENGVIRSVHEPFSKDGSLAALGDCIVKTAGVENSILKESG
PARVESODAAVKGILANEIKAGDVVIRYRGPGKGMOMELPTSYLKSGLKAC
ALITDGRSGGSLSTGHVSPPEAANGTIGLVEGDMIDIDIPNPTISLRVDAELA
ARRTQDAKQKRPVQKRRRTYTKAKVAATISADNGAVNLDGR"
repeat_region
4343: .4400
/Note="Sm-5 OR SMC04615
REPEAT SM-5
Predicted by Homology"
/evidence=not_experimental
/complement(4470: .4874)
/gene="SMC04046"
/complement(4470: .4874)
/gene="SMC04046"
/function="miscellaneous: hypothetical/global homology"
/Note="Product confidence: hypothetical
Gene name confidence: hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC47408.1"
/db_xref="GI:15075854"
/db_xref="SPTREMBL:Q92M27"
/translation="MKTRKMSCHGRIRREPVLDLFGTSRCNCSYCSKLRWGASV
KPPDRMLCEPAGIGDQFCMSCHHPRCTGCTPTPGHGVIEIGGAPFSINACID
GVDAEPALPITGMDIHNMMNAPATRM"
4979: .5024
/Note="REP (repetitive extragenic palindromic) element;
Predicted by Homology"
/evidence=not_experimental
5027: .5084
/Note="Sm-4 OR SMC04584
REPEAT SM-4
Predicted by Homology"
/evidence=not_experimental
5094: .5542
/Note="Sm-2 OR SMC04873
REPEAT SM-2
Predicted by Homology"
/evidence=not_experimental
/complement(5586: .6029)
/gene="azuz OR SMC04047"
/complement(5586: .6029)
/gene="azuz OR SMC04047"
/function="small molecule metabolism; energy transfer;
electron transport"
/Note="Product confidence: probable
Gene name confidence: putative
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE PSEUDODURIN (BLUE COPPER PROTEIN)"
/protein_id="CAC47408.1"
/db_xref="GI:15075855"
/db_xref="SPTREMBL:Q92M26"
/translation="MKTRKMLAMICATISAGQAAEERYEMLNKADGRVAFEPAY
IRAPGDTVTVAKDGCHNSALMGKAPGEGAEIWKGINETITVTLSPGVYMQCAP
HVGMGMAIIVGEPANLEAVKIKYPKSKAAAEKIPAEIESG"
complement(6076: .6417)
/gene="SMC04048"
/complement(6076: .6417)
/gene="SMC04048"

```

```

/Function="small molecule metabolism; energy transfer;
electron transport"
/Note="Product confidence: putative
Gene name confidence: hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PUTATIVE CYTOCHROME C PROTEIN"
/protein_id="CAC47410.1"
/db_xref="GI:15075856"
/db_xref="SPTREMBL:Q92M25"
/translation="MTHASSRRIAALVLCVITGLSPRMASAEEDKLAIGREIFLERSE

Query Match 100.0%; Score 15; DB 1; Length 333800;
Best Local Similarity 100.0%; Pred No. 8.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
Db 288772 ACCTACGGCATGCAC 288786

RESULT 12
HSIGVDJ2 787 bp mRNA linear PRI 10-AUG-1995
LOCUS H.sapiens rearranged Igg VHIII-D-JH-CH3 region.
DEFINITION X81696.1 GI:940516
ACCESSION X81696
VERSION 1
KEYWORDS constant region; immunoglobulin heavy chain; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 787)
AUTHORS Khamilich, A.A., Aoucturier, P., Preud'homme, J.L. and Cogne, M.
TITLE Structure of abnormal heavy chains in human heavy-chain-deposition
disease
JOURNAL Eur. J. Biochem. 229 (1), 54-60 (1995)
MEDLINE 95262687
PubMed 7744049
REFERENCE 2 (bases 1 to 787)
AUTHORS Khamilich, A.A.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1994) A.A. Khamilich, CNRS URA 1172,
LAB IMMUN AND GENETICS, IBMG FACULTY OF SCIENCES, F-86022 POITIERS
CEDEX, FRANCE

FEATURES
Source Location/Qualifiers
1..787
/organism="Homo sapiens"
/isolate="patient THR"
/db_xref="taxon:9606"
/chromosome="2"
/cell_type="plasma cell"
/tissue_type="bone marrow"
/rearranged 1..57
/product="immunoglobulin heavy chain"
/Note="gamma isotype"
58..427
/product="immunoglobulin heavy chain V-D-J"
/Note="gamma isotype"
428..750
/product="immunoglobulin heavy chain C3"
/Note="gamma isotype"
171 a 233 c 225 g 158 t
BASE COUNT 171 a 233 c 225 g 158 t
ORIGIN
Query Match 93.3%; Score 14; DB 9; Length 787;
Best Local Similarity 100.0%; Pred. No. 2e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 CCTAGGCATGCAC 15
 Db 149 CCTAGGCATGCAC 162

RESULT 13

AC010298

LOCUS 38891 bp DNA linear HTG 26-JAN-2000
 DEFINITION Homo sapiens chromosome 5 clone CTB-163N20, LOW-PASS SEQUENCE
 SAMPLING.

AC010298

AC010298.2 GI:6758851
 HTG: HTGS_PHASE0.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

Unpublished
 2 (bases 1 to 38891)
 DOE Joint Genome Institute.
 Sequencing of Human Chromosome 5
 Direct Submission
 Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Jan 26, 2000 this sequence version replaced gi:5882635.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

***** NOTE: This record contains 75 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 710: contig of 710 bp in length
 * gap of unknown length
 * 711 1503: contig of 793 bp in length
 * gap of unknown length
 * 1504 2284: contig of 781 bp in length
 * gap of unknown length
 * 2285 3002: contig of 718 bp in length
 * gap of unknown length
 * 3003 3166: contig of 164 bp in length
 * gap of unknown length
 * 3167 3745: contig of 579 bp in length
 * gap of unknown length
 * 3746 4319: contig of 574 bp in length
 * gap of unknown length
 * 4320 5096: contig of 777 bp in length
 * gap of unknown length
 * 5097 5814: contig of 718 bp in length
 * gap of unknown length
 * 5815 6228: contig of 414 bp in length
 * gap of unknown length
 * 6229 6819: contig of 591 bp in length
 * gap of unknown length
 * 6820 7319: contig of 500 bp in length
 * gap of unknown length
 * 7320 7678: contig of 359 bp in length
 * gap of unknown length
 * 7679 8199: contig of 521 bp in length
 * gap of unknown length
 * 8200 8917: contig of 718 bp in length
 * gap of unknown length

8918 9110: contig of 193 bp in length
 * gap of unknown length
 * 9111 9188: contig of 78 bp in length
 * gap of unknown length
 * 9189 9253: contig of 65 bp in length
 * gap of unknown length
 * 9254 9892: contig of 639 bp in length
 * gap of unknown length
 * 9893 10777: contig of 885 bp in length
 * gap of unknown length
 * 10778 11040: contig of 263 bp in length
 * gap of unknown length
 * 11041 11237: contig of 197 bp in length
 * gap of unknown length
 * 11238 11664: contig of 427 bp in length
 * gap of unknown length
 * 11665 12907: contig of 1243 bp in length
 * gap of unknown length
 * 12908 13764: contig of 857 bp in length
 * gap of unknown length
 * 13765 14507: contig of 743 bp in length
 * gap of unknown length
 * 14508 15239: contig of 732 bp in length
 * gap of unknown length
 * 15240 15636: contig of 397 bp in length
 * gap of unknown length
 * 15637 16504: contig of 868 bp in length
 * gap of unknown length
 * 16505 17834: contig of 1330 bp in length
 * gap of unknown length
 * 17835 17914: contig of 80 bp in length
 * gap of unknown length
 * 17915 18752: contig of 838 bp in length
 * gap of unknown length
 * 18753 18847: contig of 95 bp in length
 * gap of unknown length
 * 18848 19088: contig of 241 bp in length
 * gap of unknown length
 * 19089 19246: contig of 158 bp in length
 * gap of unknown length
 * 19247 19367: contig of 121 bp in length
 * gap of unknown length
 * 19368 19483: contig of 116 bp in length
 * gap of unknown length
 * 19484 19786: contig of 303 bp in length
 * gap of unknown length
 * 19787 20607: contig of 821 bp in length
 * gap of unknown length
 * 20608 20761: contig of 154 bp in length
 * gap of unknown length
 * 20762 20885: contig of 124 bp in length
 * gap of unknown length
 * 20886 20990: contig of 105 bp in length
 * gap of unknown length
 * 20991 21117: contig of 127 bp in length
 * gap of unknown length
 * 21118 22110: contig of 993 bp in length
 * gap of unknown length
 * 22111 23087: contig of 977 bp in length
 * gap of unknown length
 * 23088 24048: contig of 961 bp in length
 * gap of unknown length
 * 24049 24118: contig of 70 bp in length
 * gap of unknown length
 * 24119 24158: contig of 40 bp in length
 * gap of unknown length
 * 24159 24223: contig of 65 bp in length
 * gap of unknown length
 * 24224 24354: contig of 131 bp in length
 * gap of unknown length
 * 24355 24400: contig of 46 bp in length
 * gap of unknown length
 * 24401 24681: contig of 281 bp in length
 * gap of unknown length

```

* * * gap of unknown length
* 24682 25168: contig of 487 bp in length
* * * gap of unknown length
* 25169 25245: contig of 77 bp in length
* * * gap of unknown length
* 25246 25500: contig of 255 bp in length
* * * gap of unknown length
* 25501 26092: contig of 592 bp in length
* * * gap of unknown length
* 26093 26876: contig of 784 bp in length
* * * gap of unknown length
* 26877 27653: contig of 777 bp in length
* * * gap of unknown length
* 27654 27705: contig of 52 bp in length
* * * gap of unknown length
* 27706 27772: contig of 67 bp in length
* * * gap of unknown length
* 27773 29289: contig of 1517 bp in length
* * * gap of unknown length
* 29290 30054: contig of 765 bp in length
* * * gap of unknown length
* 30055 31037: contig of 983 bp in length
* * * gap of unknown length
* 31038 32220: contig of 1183 bp in length
* * * gap of unknown length
* 32221 33342: contig of 1122 bp in length
* * * gap of unknown length
* 33343 33411: contig of 69 bp in length
* * * gap of unknown length
* 33412 33451: contig of 40 bp in length
* * * gap of unknown length
* 33452 33554: contig of 103 bp in length
* * * gap of unknown length
* 33555 33607: contig of 53 bp in length
* * * gap of unknown length
* 33608 34052: contig of 445 bp in length
* * * gap of unknown length
* 34053 34494: contig of 442 bp in length
* * * gap of unknown length
* 34495 35154: contig of 660 bp in length
* * * gap of unknown length
* 35155 36132: contig of 978 bp in length
* * * gap of unknown length
* 36133 38110: contig of 1978 bp in length
* * * gap of unknown length
* 38111 38891: contig of 781 bp in length.
* * * Location/Qualifiers
* * * 1.38891
* * * /organism="Homo sapiens"
* * * /db_xref="taxon:9606"
* * * /clone="CTB-163N20"
* * * /base_count 10411 a 9018 c 8505 g 10733 t 224 others
* * *
* * * Query Match 93.3%; Score 14; DB 2; Length 38891;
* * * Best Local Similarity 100.0%; Pred. No. 5.3e+03;
* * * Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
* * *
* * * 1 ACCTACGCATGCA 14
* * * ||||||||||||
* * * Db 34462 ACCTACGCATGCA 34475

```

```

* * * tritose phosphate isomerase (tim), TATA box binding protein (TF2D),
* * * archael family B DNA polymerase B (polI), dCMP deaminase (dcdt),
* * * ATP-dependent RNA helicase, hypothetical protein 03, menaquinone
* * * biosynthesis protein (mena), and site-specific DNA
* * * methyltransferase genes, complete cds; and unknown genes.
* * * AF083072.1 GI:3599393
* * *
* * * Cenarchaeum symbiosum.
* * * Cenarchaeum symbiosum.
* * * SOURCE
* * * ORGANISM
* * * REFERENCE
* * * 1 (bases 1 to 42432)
* * * Schleper, C., Delong, E.F., Preston, C.M., Feldman, R.A., Wu, K.Y. and
* * * Swanson, R.V.
* * * Genomic analysis reveals chromosomal variation in natural
* * * populations of the uncultured psychrophilic archaean Cenarchaeum
* * * symbiosum
* * * JOURNAL
* * * U. Bacteriol. 180 (19), 5003-5009 (1998)
* * * MEDLINE
* * * 96422450
* * * PUBMED
* * * 9748430
* * * REFERENCE
* * * 2 (bases 1 to 42432)
* * * Feldman, R.A.
* * * Direct Submission
* * * Submitted (07-AUG-1998) Genomics, Diversa Corp., 10665 Sorrento
* * * Valley Rd., San Diego, CA 92121, USA
* * * FEATURES
* * * source
* * * 1.42432
* * * /organism="Cenarchaeum symbiosum"
* * * /strain="B"
* * * /db_xref="taxon:46770"
* * * /clone="6G45"
* * * /note="symbolic of the marine sponge Axinella mexicana"
* * * 3.10421
* * * /codon_start=1
* * * /transl_table=11
* * * /product="unknown"
* * * /protein_id="AAC62699.1"
* * * /db_xref="GI:3599394"
* * * /translation="MPAPGEGSLGGVAISDDGRVYATIGRDLITVYRTMPNPPDIA
* * * SAALGOSFSLPGISLPAPGAPGIDISDGRILVYDNGVYRDEAPRLGGT
* * * FGSSVYGVSAAPRCVYVAPGSLVLSDDSGRLVYDNGVYRDEAPRLGGT
* * * SDDGSEVGVAGFAGGLHMYVAGNDTRVYOTPAGTIOIDBAAGPRLISAVLDKSGTL
* * * RAAPDGVVDGVSVPGLHMYVAGNDTRVYOTPAGTIOIDBAAGPRLISAVLDKSGTL
* * * YGQSLHVPAAALAGTGGGPPVDFSGSLSLVSHNEPFGGEEMARKERDRAVLT
* * * VTAGSGQMHGAGAGNITWYDLSTPHDITGAGVAGSDIIPAVSGARNVPSITGIAF
* * * SDGMRLEAFANRGDRIPMYOLDSPYDIGSLEGTLPFGSGIARSDGTIMFAL
* * * TENAIRYDLEGPYDIRGAGNAGYOLDIPLHGLFLTSGVHSPDGRMEVGGI
* * * SDAEDNANDVNVNIMHPRDITSPPDVITAEVDPVYVSTGRAGDLESLSPGCR
* * * LITLSRRVSSSEYTTTRAOYWLPERYDTPPIHYVSPNASOGNLAADRYGNAPFSD
* * * TRLLVTGHGQTNKLEHLNPPFDVGTAVFHDHGERPGRPASITENASGISLSDSRM
* * * FLSDRGKGLISQYTLVAPEDVEFASVSDGQIDVADQADLPGLAFSPGTRLEWVG
* * * GMDRSVAMYSLNTPFDLGAENASFCVSDGSDPLGIFAGNGKTMILADITGVYAG
* * * YDCAPIYDISGPAVSGIFDAGSIIRDVAAGGSMFLIEGTGRVYBHRGIPVVSAL
* * * DGPALVSAADARVAGAEVLEFRAVVGIGDPGGRVIVAAQGLPGLVIVSDVIGED
* * * PGVAPRSTDAEVLAISGAEPLSVAGRGARVAGAGTTPSOIGNATETVSLNPTL
* * * DEGTTLTGAFNSADGTVFELSDGPTGRVYPSLNTIPDISAARGFVTVYPCVSDIA
* * * FSADGRMLVADDETGTHRYLARSVEYIGDITKSLGGEVEVFSAPRVQDLAGAF
* * * SHDGMIMLAAGSGSVHRSLSPPVAVSAGKEETMIGESPGLFESGRLRMFLSTENGID
* * * AGSETAAYVGLAIPYIGENAEPLPLFLVGAEEALISPGRIILVPGRGILSQVSLF
* * * STNIELCAEPGRGIDGSGEDGIYAFESPGGEGEVSLAASITADGPGIIEHLGAPFP
* * * MPAPVMEQVTLDSREGTLRYRDRITVDVTPRYKMWVDSGDSGOTLANSITLANS
* * * SNLLPFLDAAAGKISGYSPYRVTSSPFLCTDGTATPRLTGRCGVRLADIVYNSG
* * * DVPSGCIERSDDGMRFVGTIGTPTGINTFTLSAPDITLPHKSGSTNIGISVSDLA
* * * FANNNSLTVLDVADVLRVYALDDYNNVATGTTOKRITLDTTQGIIPNSIVTSPGLS
* * * OFVAYVDRIDLVLVIGSPNDISSTTEIIPYSLPPDPPTMDFTPPDGRMFLSTENGID
* * * QYLISEFAVTVTSVFLRTIPIDGCAEIRFVNDGRCLFVAGADGIIIOREHLIYPGAS
* * * TSLEIYRQDVITDGGGNGENPAAGEIRLAGFPNASDVOVSPGIEFGSGDGMFVFG
* * * AACVNERSLASPDITLIPVHVELHDIGGQVAPVLAEDGRITLILAAAGTIDEPVSA
* * * GDAYIDGEASRTOVPEPDAAGVAPAFVOPPGSSITIAAPCRIDQVIVVPEVSYA
* * * PLTRPPTGIDRAPGRMRFSTENGIDYLLSTFEDVRSKLTYYTITVYDVBGMQF
* * * ADNGRAFLFADSGGLIYNLDLDDVYALDGTITISVESPQSVWVYLETQTRAVSEL
* * * EFPPDVSSRRRAGTLIDIPQDSPRHVAVSPGKHLITNSVSGEDDTIHYSVINSNDI
* * * SSASYIGEEDIPPEVINGIDFSNNGRRMFLICNGPQYOVYHDMYMGSTRYDSSSL

```


DTYAIPGVVFPAGLDFSEFDRLSMFIISTAGSVVRYGLDPEFVETMDYOESFRLLPVP
 SAADNISIDLAEFGSSGLNAVISHEDGLDVLDPYGAELDIDRLPLVGVYTF
 EESDNGROLTYGAFRDSOSSPGTLFAGLORLELGPDLASAVFAOSGIDEPFNG
 MRANGLAGLHAPPDGGSTILFRAGNAERVYISIDMSHDLTLSPRESKPPVGGSTPN
 ITRDMSIDPGMFLYLQGDVDLMTNLDTSYSLDAFAVAGTLDPEDVTPGICISRD
 GTSLEMTGEDVDHIEHYALNEPMDIRNA1LNGSLISAVNGAPRGDLSDEGTAHTM
 RGDRDPTGPAVLNHLIPGOYSLTLDAPAFAPVEEGA PGDLAFSDGMRQFVAVGN
 NHLROYNLSPYDTENAHEFISTDLTADRGPTGLVFSDENDFFSTGARAFVROFTT
 NBPVASTTISDNGLYKVSVDLPSPGIRFPDGMKMFISQGETAMITYOYSLSPDPT
 SCVAVDREYIACGLFRNAGLSYGLNEPSPGDFSEDMELVYTGSGLVHRYFLPSPY
 GLEDAYGGSFHTERESTPLGVYVNGDMFVAGDSTSLKYSLA0VYGNITTHADTR
 AGIADRAELVEFGAMADTRAEILDGADVHVKSVDVFPISGITYRALYPEDAAILD
 DGMATNHNRYVITVHDITEGAPSIHDEPIAGVIALGPMDI1AVDILHRLVASLS
 GSDSPASDASGVASERNVADRGVEERLGHVSLDADRPADNMADDTDSAGVY
 RSDPGSPASDSRSGALGRMADRPADMDMTDSAGVYDRSDPGP1ISDPSALGIA
 RNAADRPAVDDMDTGSAGVYRSDPGPALSIDRSALGLARMAADRPADMDMTGSE
 STSRUGVDRPEIVERHSLASVYLSGGSPSVAGCHVDESGRDCGRDREIDTRIY
 I1TYSRGAADAPREVEDAMETSGVYAYSAGAADPAVEDAMETSGVYVPRSTMDAPT
 VADHSLARTASISDEGSPTEFAERADTVGIDIDEVDAFTVADHSLARAASISGDS
 PTEFAERADTVGIDIDEVDAFAVARLLAVLGLQADSPGVMDVIGIDHSEISGDPV
 EPRVYRGGGGGGGSSNRGLPBGGGYEIDEFPRIDGLVLFNGTDVLAESGKDLI
 RYFPRGSGFNIPDMEVILFTABGSELTAVYRAGIIMGIDCGELIMDTYISGMDLI
 I7GDETYHYERLDAFNGWYISLJGDELDTGVSYSLDNHGP1LAOHRHAKHYELLIDAA
 ENRPLSVSTDPKPVDESPVQHTESLQMDPEVESEPLPMDSEPEVDLEPVQHLSLP
 MDPEPEVDLEPVQHLPEVQSGPVOGGPESVESGIAYTLMQFLSGILLDALGLADPVG
 SVQKTS"

gene

10625..11434
 /gene="hc2"
 10625..11434

CDS

/note="similar to eukaryotic histone H1 DNA binding protein"
 /codon_start=1
 /transl_table=11
 /product="histone H1 DNA binding protein"
 /protein_id="AAC62700.1"
 /db_xref="GI:3599395"

CDS

/translation="MHGIEGGRGDMSENFVACARCVTKDEMXYVDGRVFNHECH
 ARHGQIRFPNPEVEQVRAELKYDLIOMNRNOIAENMRASGDGVHSSASAEQHR
 AEKVLQYOMRNOIAENMRNAPGKPARAKAGATATARKSKTKTGRKPARAGKAG
 ARKTKTVKRRARKTKTAKKAGKAGKAGKARRTVYVHKKIGRRKTKTARRAGKSTV
 RKSTVKTVYRVTGKKAVYRRKSTYKTRARARAGKTKTGRRARARAGARR"
 11478..13046
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAC62701.1"
 /db_xref="GI:3599396"

gene

CDS

/gene="syk"
 /codon_start=1
 /transl_table=11
 /product="lysoyl RNA synthetase"
 /protein_id="AAC62702.1"
 /db_xref="GI:3599397"
 /translation="METIGKGTWIDKLAAHLYVEREALGDTMFINVESGLGSGIHP
 MGSIGDAVRAVGVGLAVGDMGHSFRLAIFDRLDGLKVEGMPSSIEHTIAPVSAI
 PDYQCHDSYGMHNSGLLEGLDALGIEVDADDTYRGLLAEOHRLITLSVYIG
 EKIAENYGEKFRSSILYPAVCEQCGKMAEVEYLADSKVYRGGDAVEGGRIA
 GCGHEGADTGGAGKRLAMKVEFAARQWADVEFAAGKIDMSVIRINDVSEILSS
 PHPHITRYEMFLDKGKKIKSSGNTVTPQKMLRYGPQSLILLMYKRITGANELGLE

CDS

DVPSLMDVEYDLOREYFAGGGRGKAREANRGLFEYTNLLEAOEGPRPHAGYRLVE
 LSRLEFRNRERLYKVKLVEYGVIDGPPGIEURLALAGNADDMYSAERTVEVLDAT
 RGALSLELAEMYSAPEGGLQDVIYGVAKSHGVPDRFFKALYR1ILDASSGRIGPFI
 EDIGREKVACIMRGL"
 complement(117329..18213)
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein 01"
 /protein_id="AAC62703.1"
 /db_xref="GI:3599398"

rRNA

/translation="MESAGGAPGVVLDYLSKLOOYSGRDTLLYATNMWDEPHHPN
 EALLTNGDLYFRMMARDLTKKIDLLHSRGSABSAESIYVYIAKXVDIVYIIPY
 AAMSAA5LACASNSLVMGKHS1GPDADPQIFPTK1GMOINSQILLDLOVOYYS
 EKHGRUGAM1PL1GQYPPGLVQKCISSQKLAELVOKWLEDMHFAGESDAERKSKI
 SGM1ASPGKYYSHGRYISRECRNGIGLKITDLEADQFODTLVSHAAD1LSQFPI
 NK1IANH1GNSVISKPS"
 18645..21639

rRNA

/product="23S ribosomal RNA"
 21771..23243
 /product="16S ribosomal RNA"
 23558..24862

gene

/gene="gsat"
 23558..24862
 /gene="gsat"
 /function="heme biosynthesis"
 /codon_start=1
 /transl_table=11
 /product="glutamate-1-semialdehyde aminotransferase"
 /protein_id="AAC62704.1"
 /db_xref="GI:3599399"

CDS

CDS

/translation="MDLEREYRATRGSGAIFARSRRYHUGVSHNIRYEPYFVR
 SARCNHIVDVGKRTTDMGHSULIGHAFQYRSVAGSOLRGWTHGTANPTML
 SEIIRGAVAKAEKIRVYVSGTEAVMYARMARARTGKVKIACVGGWGHYASGLIKSV
 NMPYVPESSGILVDEHETVSIPYNNLEBSLEALRRAGDILACVIEPMLGGCGIPAE
 PDYLRGIDQEFVSKALG1LDEIYTGFFDFGCAVYKMGIDLPDVVALGKIVGGGPIG
 VYCGKDEVMCTSDNGAHARTERAY1GGGTFGANPATMTAGAAALGALREERGTLYPR1
 NSMGDDARARLSRIFDGRGSAVYAGRSGLPMTHTPDGARRISSAADAAACDVHLHRYH
 LDMITRQDIFFLPGKLGIAISAHSRADLGAMYSASERAGGL"
 24913..25728
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein 02"
 /protein_id="AAC62705.1"
 /db_xref="GI:3599400"
 /translation="MILFGSKSPSLLRQADLLCSGNKYKAANGLYSRILKDDPONM
 VIORGLALNRIIRYSDAITCDLLELDDSDAPYNNKALTAQAEGLGTASALENYGR
 ALEASPRYAPAYFNRAVLLDRIGHEHDALPDIDKATRLDRKANPREFYGIIVLKMGR

Query Match 93.3% Score 14; DB 1; Length 42432;

Best Local Similarity 100.0%; Pred. No. 5.1e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15

Db 13377 CCTACGGCATGCAC 13390

RESULT 15

AL357123 72354 bp DNA linear PRI 09-MAR-2001

LOCUS Human DNA sequence from clone RP11-287E6 on chromosome Xq26.3-27.3,

DEFINITION complete sequence.

ACCESSION AL357123

VERSION AL357123.11 GI:13274730

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS 1 (bases 1 to 72354)

TITLE Direct Submission

JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

COMMENT

CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi11324914.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
RP11-287B6 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: PBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-287B6 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP6-27P15 is at 72255 in this sequence.
The true right end of clone RP11-65G20 is at 100 in this sequence.

FEATURES

source
1..72354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q26.3-27.3"
/clone="RP11-287B6"
/clone_lib="RPCI-11.1"
complement(1..100)
/note="match: STS: Em:G60993
match: GSS: Em:AQ200790"
complement(916..1482)
/note="match: STS: Em:HSC19C12"
2401..2612
/note="L2 repeat: matches 2482..2669 of consensus"
2762..2804
/note="MERS repeat: matches 7..49 of consensus"
2783..3166
/note="match: GSS: Em:AQ033161"
2909..3008
/note="L1ME1 repeat: matches 5840..5947 of consensus"
4413..4608
/note="MLT1H repeat: matches 345..545 of consensus"
4774..4900
/note="MLT1G repeat: matches 43..177 of consensus"
5067..5129
/note="L2 repeat: matches 1919..1961 of consensus"
7760..7958
/note="L2 repeat: matches 2565..2750 of consensus"
7996..8111
/note="L2 repeat: matches 2575..2698 of consensus"
9225..9386
/note="L1MC5 repeat: matches 7774..7931 of consensus"
9537..10038
/note="L1ME1 repeat: matches 5431..5954 of consensus"
10762..10832
/note="L2 repeat: matches 1728..1798 of consensus"
11832..12035
/note="MIR repeat: matches 34..250 of consensus"

repeat_region 12703..13123
/note="HSMAR1 repeat: matches 1..439 of consensus"
repeat_region 13156..13426
/note="L1M4 repeat: matches 3062..3342 of consensus"
repeat_region 13465..14328
/note="L1M4 repeat: matches 2128..2968 of consensus"
repeat_region 14546..15060
/note="L1MC4 repeat: matches 7285..7835 of consensus"
repeat_region 15550..15677
/note="HAL1 repeat: matches 173..312 of consensus"
repeat_region 15763..16499
/note="L1PA10 repeat: matches 5404..6153 of consensus"
repeat_region 16490..16643
/note="HAL1 repeat: matches 408..759 of consensus"
repeat_region 16966..17089
/note="62 copies 2 mer tt 71% conserved"
repeat_region 17051..17102
/note="13 copies 4 mer ctcc 92% conserved"
repeat_region 17115..17414
/note="ALUSx repeat: matches 1..301 of consensus"
repeat_region 17578..18034
/note="L1PA13 repeat: matches 5695..6152 of consensus"
repeat_region 18379..18691
/note="ALUSx repeat: matches 1..310 of consensus"
repeat_region 20868..21162
/note="ALUSg repeat: matches 1..296 of consensus"
repeat_region 21417..21600
/note="MIR repeat: matches 35..212 of consensus"
repeat_region 22265..22415
/note="MIR repeat: matches 2..143 of consensus"
repeat_region 22581..22664
/note="21 copies 4 mer atag 78% conserved"
repeat_region 24520..24721
/note="MERSB repeat: matches 6..178 of consensus"
repeat_region 25456..27803
/note="L1 repeat: matches 2387..4703 of consensus"
repeat_region 28250..28835
/note="L1 repeat: matches 4117..4755 of consensus"
repeat_region 29493..29573
/note="L1 repeat: matches 4023..4106 of consensus"
repeat_region 29574..29870
/note="ALUSg repeat: matches 1..297 of consensus"
repeat_region 29871..29964
/note="L1 repeat: matches 3939..4023 of consensus"
repeat_region 30234..30588
/note="THE1A repeat: matches 1..354 of consensus"
repeat_region 31405..31491
/note="MLT1I repeat: matches 1..85 of consensus"
repeat_region 32756..33050
/note="ALUSg repeat: matches 1..295 of consensus"
repeat_region 33150..33283
/note="L2 repeat: matches 808..951 of consensus"
repeat_region 33470..33821
/note="MLT1F repeat: matches 201..539 of consensus"
repeat_region 33881..34193
/note="ALUSx repeat: matches 1..312 of consensus"
repeat_region 34199..34368
/note="MLT1F repeat: matches 10..172 of consensus"
repeat_region 34366..34822
/note="MLT1A1 repeat: matches 1..365 of consensus"
repeat_region 34824..34960
/note="L1M3 repeat: matches 5464..5597 of consensus"
repeat_region 34958..35119
/note="L1M4 repeat: matches 3673..3836 of consensus"
repeat_region 35159..35248
/note="L2 repeat: matches 1249..1337 of consensus"
repeat_region 35249..35553
/note="ALUSx repeat: matches 1..305 of consensus"
repeat_region 35554..35948
/note="L2 repeat: matches 1337..1739 of consensus"
repeat_region 35957..36082
/note="L2 repeat: matches 1945..2077 of consensus"
repeat_region 36358..37198

```

repeat_region /note="LIMB6 repeat: matches 5318. .6174 of consensus"
37209. .37447 /note="LIPa15 repeat: matches 5918. .6157 of consensus"
repeat_region 37475. .37531 /note="LIP repeat: matches 2395. .2451 of consensus"
repeat_region 37571. .39494 /note="LIPa13 repeat: matches 523. .2465 of consensus"
repeat_region 39502. .40031 /note="LIM4 repeat: matches 4793. .5331 of consensus"
repeat_region 40379. .41594 /note="LIM4 repeat: matches 2614. .3869 of consensus"
repeat_region 41599. .41684 /note="LIM4 repeat: matches 1225. .1319 of consensus"
repeat_region 41685. .41978 /note="LIM4 repeat: matches 1. .306 of consensus"
repeat_region 41979. .41992 /note="LIM4 repeat: matches 1213. .1225 of consensus"
repeat_region 41998. .42194 /note="LIM4 repeat: matches 116. .313 of consensus"
repeat_region 42199. .43409 /note="LIM4 repeat: matches -12. .1191 of consensus"
repeat_region 43430. .43525 /note="LIM4 repeat: matches 2399. .2495 of consensus"
repeat_region 43606. .43736 /note="LIM4 repeat: matches 2614. .2750 of consensus"
repeat_region 44810. .46062 /note="LIM4 repeat: matches 1467. .2749 of consensus"
repeat_region 46166. .46419 /note="LIM4 repeat: matches 2. .255 of consensus"
repeat_region 46662. .46809 /note="LIM4 repeat: matches 4611. .4760 of consensus"
repeat_region 46910. .48311 /note="LIM4 repeat: matches 4755. .6158 of consensus"
repeat_region 48340. .48375 /note="LIM4 repeat: matches 5386. .5419 of consensus"
repeat_region 48376. .48668 /note="LIM4 repeat: matches 1. .295 of consensus"
repeat_region 48669. .49381 /note="LIM4 repeat: matches 5419. .6164 of consensus"
repeat_region 49857. .49964 /note="LIM4 repeat: matches 5419. .6164 of consensus"

```

```

Query Match 93.38; Score 14; DB 9; Length 72354;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 CCTACGCGCATGCAC 15
   |||||||||||||
Db 11377 CCTACGCGCATGCAC 11390

```

```

Search completed: June 26, 2003, 03:47:32
Job time : 428.6 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-25
Perfect score: 15
Sequence: 1 acctacgcacatgcac 15

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	20	AA32399
2	15	100.0	115	24	ABK71361
3	15	100.0	115	24	ABA04558
4	15	100.0	351	24	ABK71358
5	15	100.0	351	24	ABA04555
6	15	100.0	432	24	ABK71369
7	15	100.0	432	24	ABA04563
8	15	100.0	717	20	AAK17988
9	15	100.0	823	24	ABK71387

Result ID	Score	Query Match	Length	DB ID	Description	ALIGNMENTS
10	15	100.0	2908	24	ABL56609	Nucleotide sequenc
11	14	93.3	1575	21	AA55191	Genarchaeum symbio
12	14	93.3	42432	21	AA55187	Mouse spliced tran
13	13.4	89.3	65	24	ABN57826	Corn tassell-derive
14	13.4	89.3	292	24	ABL71855	Human anti-Rh(D) c
15	13.4	89.3	372	22	AAH68618	Human anti-Rh(D) c
16	13.4	89.3	375	22	AAH68623	Human anti-Rh(D) c
17	13.4	89.3	375	22	AAH68638	Human anti-Rh(D) c
18	13.4	89.3	429	21	AAAI3930	Human PTHrP monocl
19	13.4	89.3	429	21	AAAI3932	Human PTHrP monocl
20	13.4	89.3	429	21	AAAI3933	Human PTHrP monocl
21	13.4	89.3	429	21	AAAI3935	Human PTHrP monocl
22	13.4	89.3	429	21	AAAI3937	Human PTHrP monocl
23	13.4	89.3	471	20	AAK91539	Porphyromonas ging
24	13.4	89.3	714	24	ABK36221	CDNA sequence #612
25	13.4	89.3	936	22	AAH43565	Human DHDR-2 codin
26	13.4	89.3	1157	21	AAK39072	Human secreted pro
27	13.4	89.3	1320	20	AAK75798	H. pylori outer me
28	13.4	89.3	1329	18	AAH68267	H. pylori cytoplas
29	13.4	89.3	1379	22	AAH43561	Human DHDR-2 CDNA
30	13.4	89.3	1612	23	AAK79266	DNA encoding novel
31	13.4	89.3	1689	22	AAK72302	Corynebacterium q1
32	13.4	89.3	1743	22	AAH66792	C glutamicum codin
33	13.4	89.3	2253	24	ABK84330	Human cDNA differe
34	13.4	89.3	2253	24	ABK84330	Kidney cancer rela
35	13.4	89.3	2472	22	ABA89120	Escherichia coli p
36	13.4	89.3	2939	21	AAK86410	Wheat starch synth
37	13.4	89.3	2939	21	AAK86433	Wheat starch synth
38	13.4	89.3	3054	23	ABL04210	Drosophila melanog
39	13.4	89.3	3735	19	AAV69425	Enrichia sp. exte
40	13.4	89.3	3735	21	AAK89990	Enrichia sp. exte
41	13.4	89.3	3735	24	AAK89990	DNA encoding human
42	13.4	89.3	3735	24	AAK89990	DNA encoding human
43	13.4	89.3	3735	24	AAK89990	DNA encoding human
44	13.4	89.3	4916	23	ABL18373	Drosophila melanog
45	13.4	89.3	6594	23	ABL15639	Drosophila melanog

Result ID	Score	Query Match	Length	DB ID	Description	ALIGNMENTS
1	15	100.0	2908	24	ABL56609	Nucleotide sequenc
2	14	93.3	1575	21	AA55191	Genarchaeum symbio
3	14	93.3	42432	21	AA55187	Mouse spliced tran
4	13.4	89.3	65	24	ABN57826	Corn tassell-derive
5	13.4	89.3	292	24	ABL71855	Human anti-Rh(D) c
6	13.4	89.3	372	22	AAH68618	Human anti-Rh(D) c
7	13.4	89.3	375	22	AAH68623	Human anti-Rh(D) c
8	13.4	89.3	375	22	AAH68638	Human anti-Rh(D) c
9	13.4	89.3	429	21	AAAI3930	Human PTHrP monocl
10	13.4	89.3	429	21	AAAI3932	Human PTHrP monocl
11	13.4	89.3	429	21	AAAI3933	Human PTHrP monocl
12	13.4	89.3	429	21	AAAI3935	Human PTHrP monocl
13	13.4	89.3	429	21	AAAI3937	Human PTHrP monocl
14	13.4	89.3	471	20	AAK91539	Porphyromonas ging
15	13.4	89.3	714	24	ABK36221	CDNA sequence #612
16	13.4	89.3	936	22	AAH43565	Human DHDR-2 codin
17	13.4	89.3	1157	21	AAK39072	Human secreted pro
18	13.4	89.3	1320	20	AAK75798	H. pylori outer me
19	13.4	89.3	1329	18	AAH68267	H. pylori cytoplas
20	13.4	89.3	1379	22	AAH43561	Human DHDR-2 CDNA
21	13.4	89.3	1612	23	AAK79266	DNA encoding novel
22	13.4	89.3	1689	22	AAK72302	Corynebacterium q1
23	13.4	89.3	1743	22	AAH66792	C glutamicum codin
24	13.4	89.3	2253	24	ABK84330	Human cDNA differe
25	13.4	89.3	2253	24	ABK84330	Kidney cancer rela
26	13.4	89.3	2472	22	ABA89120	Escherichia coli p
27	13.4	89.3	2939	21	AAK86410	Wheat starch synth
28	13.4	89.3	2939	21	AAK86433	Wheat starch synth
29	13.4	89.3	3054	23	ABL04210	Drosophila melanog
30	13.4	89.3	3735	19	AAV69425	Enrichia sp. exte
31	13.4	89.3	3735	21	AAK89990	Enrichia sp. exte
32	13.4	89.3	3735	24	AAK89990	DNA encoding human
33	13.4	89.3	3735	24	AAK89990	DNA encoding human
34	13.4	89.3	3735	24	AAK89990	DNA encoding human
35	13.4	89.3	3735	24	AAK89990	DNA encoding human
36	13.4	89.3	4916	23	ABL18373	Drosophila melanog
37	13.4	89.3	6594	23	ABL15639	Drosophila melanog

DR	P-PSDB; AAU06699.
XX	
PT	New thrombopoietin receptor agonist antibodies - useful for
PT	treating immunological or hematological disorders
XX	
PS	Claim 10; Page 78; 86pp; English.
XX	
CC	The invention relates to an agonist antibody (Ab) which binds to a
CC	thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
CC	be used in the same way and for the same indications as thrombopoietin
CC	(TPO). They can stimulate proliferation, differentiation or growth of
CC	megakaryocytes. They may also be able to stimulate megakaryocytes to
CC	increase platelet production. They can be used for treating
CC	immunological or hematopoietic disorders, especially thrombocytopenia.
CC	Thrombocytopenia -associated bone marrow hypoplasia (e.g. aplastic anemia
CC	following chemotherapy or bone marrow transplant) may be effectively
CC	treated with the antibody compounds as well as disorders such as
CC	disseminated intravascular coagulation (DIC), immune thrombocytopenia,
CC	(HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
CC	congenital thrombocytopenia, thrombotic thrombocytopenia and
CC	myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
CC	treatment of solid tumours or leukaemia, myeloblastic chemotherapy for
CC	autologous or allogeneic bone marrow transplant, myelodysplasia, and
CC	idiopathic aplastic anemia, congenital thrombocytopenia, and immune
CC	thrombocytopenia. The antibodies which bind to the MxkR receptor can be
CC	used for improving neuromuscular function in a patient, e.g. in muscular
CC	dystrophy. The products can also be used for detection and diagnosis. The
CC	antibodies have a longer half-life than the natural ligand for the TPO-R.
CC	Sequences AAX2387-X32413 represent DNA fragments encoding the CDR1,
CC	CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
CC	(VL) chains of antibodies Ab1 to Ab6.
SO	
XX	Sequence 15 BP; 4 A; 6 C; 3 G; 2 T; 0 other;
Query Match	100.0%; Score 15; DB 20; Length 15;
Best Local Similarity	100.0%; Pred. No. 51;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ACCTACGGCATGTCAC 15
Db	1 ACTTAGCGCATGCAC 15
RESULT 2	
ID	ABK71361/c
XX	ABK71361 standard; DNA; 115 BP.
AC	ABK71361;
DT	30-JUL-2002 (first entry)
DE	Thrombopoietin (TPO) agonist antibody associated polynucleotide #2.
XX	
KW	Modified antibody; thrombopoietin; TPO; agonist;
RN	TPO receptor; platelet reduction-associated blood disease;
RV	thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
XX	ds.
OS	Synthetic.
PN	WO200233072-A1.
PD	25-APR-2002.
PF	22-OCT-2001; 2001WO-JP09259.
PR	20-OCT-2000; 2000JP-0321821.
PA	17-APR-2001; 2001WO-JP03288.
TI	12-SEP-2001; 2001JP-0277314.
XX	
XX	(CHUS) CHUGAI SEIRAKU KK.
XX	
XX	Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

XX WPI; 2002-383513/41.
XX
XX Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
PT leukemia -
XX
XX Example 7; Page 178; 213pp; Japanese.
PS
XX The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives
CC and/or remedies for platelet reduction-associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukaemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence represents thrombopoietin (TPO) agonist antibody associated
CC polynucleotide used in the creation of the modified antibody described in
CC the invention.
XX
SQ Sequence 115 BP; 25 A; 39 C; 33 G; 18 T; 0 other;
XX

Query Match 100.0%; Score 15; DB 24; Length 115;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCAGTCAC 15
|||||
DB 68 ACCATCGGCATGCA 54 .

ABR04558/c
ID ABA04558 standard; DNA; 115 BP.
XX
AC ABA04558;
XX
DT 15-FEB-2002 (first entry)
XX
DE Oligonucleotide #2.
XX
KW Cytostatic; antiinflammatory; antianaemic; vasotropic;
KW antibody; signal transmission; cancer; inflammation; hormonal disorder;
KW leukaemia; lymphoma; aplastic anaemia; skeletal malformation; ss.
XX
OS Synthetic.
XX
PN WO200179494-A1.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-JP03288.
XX
PR 17-APR-2000; 2000JP-0115246.
PR 20-OCT-2000; 2000JP-0321821.
PR 20-OCT-2000; 2000JP-0321822.
PR 12-MAR-2001; 2001WO-JP01912.
PA
XX (CHUS) CHUGAI SEIYAKU KK.
XX
PL Fukushima N, Tsuchiya M, Oheda M, Uno S, Kikuchi Y, Ohtomo T;
DR WPI; 2002-066368/09.
XX
XX Antibodies for treatment of diseases associated with cell
PT proliferation, hormonal disorders and cytokines comprise agonist
PT activity to signal transmission across cell membranes -
XX
XX Example 7; Page 157; 173pp; Japanese.

The present invention relates to modified antibodies. Tho antibodies

CC contain two or more H chain V domains and two or more L chain V domains
 CC of a monoclonal antibody (Mab) which is capable of transmitting a signal
 CC across the cell membrane by cross-linking a cell surface molecule, where
 CC the antibodies can serve as signal transmission agonists. The antibodies
 CC are useful for treatment and prevention of a broad range of disorders in
 CC which signal transmission is implicated, such as cancer, inflammation,
 CC hormonal disorders, leukaemia, malignant lymphoma, aplastic anaemia, and
 CC skeletal malformations. The present sequence was used in an example from
 CC the present invention.

XX Sequence 115 BP; 25 A; 39 C; 33 G; 18 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 115;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
 |||
 DB 68 ACCTACGGCATGCAC 54

RESULT 4
 ABK71358
 ID ABK71358 standard; DNA: 351 BP.

AC ABK71358;

DF 30-JUL-2002 (first entry)

XX DNA encoding thrombopoietin agonist antibody associated protein #14.

DE Modified antibody: thrombopoietin; TPO; agonist;

KW TPO receptor; platelet reduction-associated blood disease;

KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;

KW gene; ds.

OS Homo sapiens.

PN WO200233072-A1.

XX 25-APR-2002.

PF 22-OCT-2001; 2001WO-JP09259.

PR 20-OCT-2000; 2000JP-0321821.

PR 17-APR-2001; 2001WO-JP03288.

PR 12-SEP-2001; 2001JP-0277314.

PA (CHUS) CHUGAI SEIYAKU KK.

PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

XX WPI: 2002-383513/41.

DR P-PSDB; ABG35320.

XX Degraded thrombopoietin agonist antibodies containing H and L chain V

XX domains of monoclonal antibody, useful in preventives and/or remedies

XX for blood diseases, thrombocytopenia following cancer chemotherapy or

XX leukaemia

XX Example 7; Page 175-176; 213pp; Japanese.

XX The invention describes a modified antibody comprising at least 2 heavy
 XX chain variable domains and 2 or more light chain variable domains of an
 XX antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
 XX the TPO receptor to crosslink. The antibodies are useful in preventives
 XX and/or remedies for platelet reduction associated blood diseases,
 XX thrombocytopenia following cancer chemotherapy or leukaemia. The
 XX antibody can act as a TPO signal transduction agonist by transducing a
 XX signal into cells by crosslinking a TPO receptor to exert TPO agonism.
 XX This sequence encodes a thrombopoietin (TPO) agonist antibody associated
 XX protein.

SQ Sequence 351 BP; 75 A; 90 C; 113 G; 73 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 351;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
 |||
 DB 91 ACCTACGGCATGCAC 105

RESULT 5
 ABA04555
 ID ABA04555 standard; DNA: 351 BP.

XX ABA04555;

DT 15-FEB-2002 (first entry)

XX Human MPL #1 coding sequence.

XX Human; cytostatic; antiinflammatory; antinaeemic; vasotropic;

KW antibody; signal transmission; cancer; inflammation; hormonal disorder;

XX Leukaemia; lymphoma; aplastic anaemia; skeletal malformation; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..351

FT /tag= a

FT /partial

FT /product= "Human MPL #1"

FT /note= "No start or stop codon given"

XX WO200179494-A1.

XX 25-OCT-2001.

PF 17-APR-2001; 2001WO-JP03288.

PR 17-APR-2000; 2000JP-0115246.

PR 20-OCT-2000; 2000JP-0321821.

PR 20-OCT-2000; 2000JP-0321822.

PR 12-MAR-2001; 2001WO-JP01912.

PA (CHUS) CHUGAI SEIYAKU KK.

PI Fukushima N, Tsuchiya M, Oheda M, Uno S, Kikuchi Y, Ohtomo T;

XX WPI: 2002-066368/09.

DR P-PSDB; AAM47640.

XX Antibodies for treatment of diseases associated with cell

XX proliferation, hormonal disorders and cytokines comprise agonist

XX activity to signal transmission across cell membranes

XX Example 7; Page 155-156; 173pp; Japanese.

XX The present invention relates to modified antibodies. The antibodies
 XX contain two or more H chain V domains and two or more L chain V domains
 XX of a monoclonal antibody (Mab) which is capable of transmitting a signal
 XX across the cell membrane by cross-linking a cell surface molecule, where
 XX the antibodies can serve as signal transmission agonists. The antibodies
 XX are useful for treatment and prevention of a broad range of disorders in
 XX which signal transmission is implicated, such as cancer, inflammation,
 XX hormonal disorders, leukaemia, malignant lymphoma, aplastic anaemia, and
 XX skeletal malformations. The present sequence was used to illustrate the
 XX present invention.

XX Sequence 351 BP; 75 A; 90 C; 113 G; 73 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 351;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
 |||||||
 DB 91 ACCTACGGCATGCAC 105

RESULT 6

ABK71369
 ID ABK71369 standard; DNA; 432 BP.

AC ABK71369;

DT 30-JUL-2002 (first entry)

DE DNA encoding thrombopoietin agonist antibody associated protein #17.

XX Modified antibody; thrombopoietin; TPO; agonist;

KM TPO receptor; platelet reduction associated blood disease;

KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;

OS Homo sapiens.

PN WO200233072-A1.

PD 25-APR-2002.

PE 22-OCT-2001; 2001WO-JP09259.

PR 20-OCT-2000; 2000JP-0321821.

PR 17-APR-2001; 2001WO-JP03288.

PR 12-SEP-2001; 2001JP-0277314.

XX (CHUS) CHUGAI SEIYAKU KK.

PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

DR WPI: 2002-383513/41.

DR P-PSDB; ABG35323.

PT Degraded thrombopoietin agonist antibodies containing H and L chain V

PT domains of monoclonal antibody, useful in preventives and/or remedies

PT for blood diseases, thrombocytopenia following cancer chemotherapy or

XX leukaemia

PS Example 7; Page 181-182; 213pp; Japanese.

CC The invention describes a modified antibody comprising at least 2 heavy

CC chain variable domains and 2 or more light chain variable domains of an

CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing

CC the TPO receptor to crosslink. The antibodies are useful in preventives

CC and/or remedies for platelet reduction-associated blood diseases,

CC thrombocytopenia following cancer chemotherapy or leukaemia. The

CC antibody can act as a TPO signal transduction agonist by transducing a

CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.

CC This sequence encodes a thrombopoietin (TPO) agonist antibody associated

XX protein.

SO Sequence 432 BP; 87 A; 106 C; 139 G; 100 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 432;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
 |||||||
 DB 159 ACCTACGGCATGCAC 173

RESULT 7

ABK04563
 ID ABK04563 standard; DNA; 432 BP.

XX ABA04563;

AC 15-FEB-2002 (first entry)

DT Human coding sequence, SEQ ID 63.

XX Human; cytostatic; antiinflammatory; antianaemic; vasotropic;

KW antibody; signal transduction; cancer; inflammation; hormonal disorder;

KW leukaemia; lymphoma; aplastic anaemia; skeletal malformation; ds.

OS Homo sapiens.

PN Key Location/Qualifiers

FT CDS 12..419

FT /tag= a

FT /partial

FT /product= "Human protein"

FT /note= "No stop codon given"

PN WO200179494-A1.

PD 25-OCT-2001.

PE 17-APR-2001; 2001WO-JP03288.

PR 17-APR-2000; 2000JP-0115246.

PR 20-OCT-2000; 2000JP-0321821.

PR 20-OCT-2000; 2000JP-0321822.

PR 12-MAR-2001; 2001WO-JP01912.

XX (CHUS) CHUGAI SEIYAKU KK.

PI Fukushima N, Tsuchiya M, Oheda M, Uno S, Kikuchi Y, Ohtomo T;

DR WPI: 2002-066368/09.

DR P-PSDB: AAM47642.

PT Antibodies for treatment of diseases associated with cell

PT proliferation, hormonal disorders and cytokines comprise agonist

PT activity to signal transduction across cell membranes -

XX Example 7; Page 159-160; 173pp; Japanese.

CC The present invention relates to modified antibodies. The antibodies

CC contain two or more H chain V domains and two or more L chain V domains

CC of a monoclonal antibody (MAB) which is capable of transmitting a signal

CC across the cell membrane by cross-linking a cell surface molecule, where

CC the antibodies can serve as signal transduction agonists. The antibodies

CC are useful for treatment and prevention of a broad range of disorders in

CC which signal transduction is implicated, such as cancer, inflammation,

CC hormonal disorders, leukaemia, malignant lymphoma, aplastic anaemia, and

CC skeletal malformations. The present sequence was used to illustrate the

XX present invention.

SO Sequence 432 BP; 87 A; 106 C; 139 G; 100 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 432;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
 |||||||
 DB 159 ACCTACGGCATGCAC 173

RESULT 8

AAK17988
 ID AAK17988 standard; DNA; 717 BP.

AC AAK17988;

DT 11-MAY-1999 (first entry)


```

XX Anti-Mpl scFv 12B5 coding sequence.
DE
XX
XX Variant; antibody; heavy chain; light chain; immunoadhesin; immunoassay;
KW diagnosis; cancer; primer; PCR; amplification; diclstonic; ss.
XX
XX Synthetic.
OS
XX WO9850431-A2.
PN
XX 12-NOV-1998.
PD
XX 30-APR-1998; 98WO-US08762.
PE
XX 24-JUN-1997; 97US-0050661.
PR 02-MAY-1997; 97US-0850058.
XX
XX (GETH ) GENENTECH INC.
PA
XX Arathoon R, Carter PJ, Merchant AM, Presta LG;
PI
XX WPI; 1999-070091/06.
DR
XX
XX Selective preparation of multispecific antibodies - with
PT heteromultimeric heavy chain and common light chain components,
PT useful for, e.g. in vivo diagnosis of cancer
XX
XX Example 4; Page -: 69pp; English.
PS
XX This sequence represents the coding sequence of the anti-Mpl scFv
CC antibody 12B5. The sequence encoding the chain is generated by a new
CC method for preparing a multispecific Ab comprising a first polypeptide
CC (PP) and at least 1 extra PP, where: (i) the first PP comprises a
CC multimerisation domain (MD) forming an interface positioned to interact
CC with an interface of a MD of the extra PP, and (ii) the first and extra
CC PPs each have a binding domain, which comprises a heavy chain and a light
CC chain, where the variable light chains of the first and extra PPs
CC comprise a common sequence. The method comprises: (a) culturing a host
CC cell comprising nucleic acid encoding the first PP and extra PP, and
CC the variable light chain, such that the nucleic acid is expressed; and
CC (b) recovering the multispecific Ab from the culture. The method prepares
CC heteromultimeric PPs, such as bispecific Abs, bispecific immunoadhesins
CC and Ab-immunoadhesin chimeras. The method allows for the enhanced
CC formation of the desired heteromultimer relative to the undesired
CC heteromultimers and homomultimers. The Abs can be used in immunoassays
CC and for the in vitro or in vivo diagnosis of various diseases, such as
CC cancer.
XX
XX Sequence 717 BP; 161 A; 188 C; 215 G; 153 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 20; Length 717;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGGCATGCAC 15
Db 91 ACCTACGGCATGCAC 105

```

```

XX Homo sapiens.
OS
XX WO200233072-A1.
PN
XX 25-APR-2002.
PD
XX 22-OCT-2001; 2001WO-JP09259.
PE
XX 20-OCT-2000; 2000JP-0321821.
PR 17-APR-2001; 2001WO-JP03288.
PR 12-SEP-2001; 2001JP-0277314.
XX
XX (CHUS ) CHUGAI SEIYAKU KK.
PA
XX Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;
PI
XX WPI; 2002-383513/41.
DR
XX P-PSDB; ABG35329.
DR
XX
XX Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
PT leukaemia -
XX
XX Example 7; Page 190-192; 213pp; Japanese.
PS
XX The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives,
CC and/or remedies for platelet reduction reduction-associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukaemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence encodes a thrombopoietin (TPO) agonist antibody associated
CC protein.
XX
XX Sequence 823 BP; 179 A; 206 C; 248 G; 190 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 24; Length 823;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGGCATGCAC 15
Db 159 ACCTACGGCATGCAC 173

```

```

RESULT 9
ABK71387
ID ABK71387 standard; DNA; 823 BP.
XX
XX ABK71387;
AC
XX
XX 30-JUL-2002 (first entry)
DT
XX
XX DNA encoding thrombopoietin agonist antibody associated protein #21.
DE
XX
XX Modified antibody; thrombopoietin; TPO; agonist;
KW TPO receptor; platelet reduction-associated blood disease;
KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
KW gene; ds.

```

```

RESULT 10
ABL56609/C
ID ABL56609 standard; cDNA; 2908 BP.
XX
XX ABL56609;
AC
XX
XX 30-JUL-2002 (first entry)
DT
XX
XX Nucleotide sequence of human dehydrogenase 12.
DE
XX
XX Human; dehydrogenase 12; cancer; HIV infection; enzyme; gene; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 937..1254
FT /*tag= a
FT /product= "dehydrogenase 12"
XX
XX CNI325967-A.
PN
XX
XX 12-DEC-2001.
XX
XX 26-MAY-2000; 2000CN-0115887.
PF

```

```

XX 26-MAY-2000; 2000CN-0115887.
PR
XX
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
PA
XX Mao Y, xie Y;
PI
XX WPI: 2002-196688/26.
DR
XX P-PSDB; ABB09889.
PT
XX Polypeptide-human dehydrogenase 12 and polynucleotide encoding it -
PS
XX Claim 6; Page 24-26 (disclosure); 32pp; Chinese.
XX
XX The present sequence encodes human dehydrogenase 12. The polypeptide
CC is used for treating diseases such as cancer and HIV infection.
XX
SQ Sequence 2908 BP; 809 A; 539 C; 614 G; 946 T; 0 other;

Query Match          100.0%; Score 15; DB 24; Length 2908;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 ACCTACGGCATGCAC 15
        |||||||
Db      1239 ACCTACGGCATGCAC 1225

RESULT 11
AAA55191
ID AAA55191 standard; DNA; 1575 BP.
AC
XX AAA55191;
XX
XX 30-AUG-2000 (first entry)
DT
XX
XX Cenarchaeum symbiosum open reading frame nucleotide sequence SEQ ID NO:9.
DE
XX
XX Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW characterisation; archae; therapeutic; industrial; laboratory; ds.
XX
XX Cenarchaeum symbiosum.
OS
XX
XX WO200018909-A2.
PN
XX
XX 06-APR-2000.
PD
XX
XX 29-SEP-1999; 99WO-US22752.
PF
XX
XX 29-SEP-1998; 98US-0102294.
PR
XX
XX (DIVE-) DIVERSA CORP.
PA
XX
XX Swanson RV, Feldman RA, Schleper C;
PI
XX
XX WPI: 2000-293148/25.
DR
XX
XX P-PSDB; AAY90916.
PT
XX
XX New nucleic acids and proteins isolated from the non-thermophilic
PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT physiology of these archae and in therapeutic, industrial or laboratory
PT techniques -
XX
XX Claim 7; Page 116-118; 210pp; English.
XX
XX AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
CC and proteins isolated from the non-thermophilic crenarchaeote
CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in
CC the present invention are useful in characterizing the physiology of
CC these archae and can be used in therapeutic, industrial or laboratory
CC techniques. AAA55227 to AAA55260 represent promoter sequences from
CC Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
CC probes used in examples from the present invention.

```

```

XX SQ Sequence 1575 BP; 356 A; 424 C; 548 G; 247 T; 0 other;
XX
XX Query Match          93.3%; Score 14; DB 21; Length 1575;
XX Best Local Similarity 100.0%; Pred. No. 2.3e+02;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 CCTACGGCATGCAC 15
        |||||||
Db      332 CCTACGGCATGCAC 345

RESULT 12
AAA55187
ID AAA55187 standard; DNA; 42432 BP.
AC
XX AAA55187;
XX
XX 30-AUG-2000 (first entry)
DT
XX
XX Cenarchaeum symbiosum nucleotide sequence variant B SEQ ID NO:2.
DE
XX
XX Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW characterisation; archae; therapeutic; industrial; laboratory; ds.
XX
XX Cenarchaeum symbiosum.
OS
XX
XX WO200018909-A2.
PN
XX
XX 06-APR-2000.
PD
XX
XX 29-SEP-1999; 99WO-US22752.
PF
XX
XX 29-SEP-1998; 98US-0102294.
PR
XX
XX (DIVE-) DIVERSA CORP.
PA
XX
XX Swanson RV, Feldman RA, Schleper C;
PI
XX
XX WPI: 2000-293148/25.
DR
XX
XX New nucleic acids and proteins isolated from the non-thermophilic
PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT physiology of these archae and in therapeutic, industrial or laboratory
PT techniques -
XX
XX Claim 1; Page 75-87; 210pp; English.
XX
XX AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
CC and proteins isolated from the non-thermophilic crenarchaeote
CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in
CC the present invention are useful in characterizing the physiology of
CC these archae and can be used in therapeutic, industrial or laboratory
CC techniques. AAA55227 to AAA55260 represent promoter sequences from
CC Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
CC probes used in examples from the present invention.
XX
XX SQ Sequence 42432 BP; 8792 A; 12248 C; 12606 G; 8786 T; 0 other;

Query Match          93.3%; Score 14; DB 21; Length 42432;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 CCTACGGCATGCAC 15
        |||||||
Db      13377 CCTACGGCATGCAC 13390

RESULT 13
ABN57826
ID ABN57826 standard; DNA; 65 BP.
XX
XX AC ABN57826;

```

XX		15-JUL-2002	(first entry)
DT			
DE		Mouse spliced transcript detection oligonucleotide SEQ ID NO:30574.	
XX			
XX		Human; mouse; rat; splice transcript; detection; RNA transcript;	
KW		splice variant; transcriptome; oligonucleotide library; ss.	
XX			
OS		Mus musculus.	
XX			
WO		WO200210449-A2.	
XX			
PN		07-FEB-2002.	
PD			
PF		20-JUL-2001; 2001WO-IB01903.	
XX			
PR		28-JUL-2000; 2000US-221607P.	
PR		02-MAY-2001; 2001US-287724P.	
XX			
PA		(COMP-) COMPUGEN INC.	
XX			
PI		Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;	
XX			
DR		WPI: 2002-257383/30.	
XX			
PT		New oligonucleotide libraries comprising oligonucleotides which	
PT		selectively hybridize to mRNAs transcribed from a transcription unit of	
PT		a genome, useful for detecting tissue-, pathology-, and	
PT		developmental-specific genes -	
PS			
PS		Example 1; SEQ ID 30574; 47pp: English.	
XX			
CC		The present invention describes oligonucleotide libraries for detecting	
CC		messenger RNAs that populate a (sub-)transcriptome, where the	
CC		(sub-)transcriptome comprises messenger RNAs transcribed from multiple	
CC		transcription units that populate a genome. The library comprises	
CC		several oligonucleotides, each capable of hybridising selectively to a	
CC		set of messenger RNAs transcribed from a given transcription unit of	
CC		the genome, which encodes one or more messenger RNA splice variants.	
CC		The oligonucleotide libraries are useful for detecting mRNAs from a	
CC		biological sample, in expression profiling studies, in qualitatively or	
CC		quantitatively characterising the corresponding transcriptome, and in	
CC		detecting RNA transcripts and splice variants of human or animal	
CC		transcriptomes. The libraries may also be used as specialised mini	
CC		libraries to detect transcripts of a sub-transcriptome under a	
CC		particular biological or pathological state, and so allowing the	
CC		detection of tissue- and pathology-specific genes such as those genes	
CC		only expressed in specific tissue under a specific pathological	
CC		condition: to detect developmental specific genes; and to detect RNA	
CC		transcripts and splice variants of a transcriptome of a patient suffering	
CC		from a particular disorder. ABN272253 to ABN59589 represent	
CC		oligonucleotide sequences from rats, humans and mice, which are used in	
CC		the exemplification of the present invention.	
CC		N.B. The sequence data for this patent did not form part of the printed	
CC		specification, but was obtained in electronic format directly from WIPO	
CC		at ftp.wipo.int/pub/published_pcl_sequences.	
XX			
SQ		Sequence 65 BP; 17 A; 20 C; 13 G; 15 T; 0 other:	
		Query Match	89.3%; Score 13.4; DB 24; Length 65;
		Best Local Similarity	93.3%; Pred. No. 4.3e+02;
		Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
OY		1 ACCTACGGCATGCAC 15	
DB		14 ACTTAGGCGATGCAC 28	
		RESULT 14	
ID		ABL/71855/c	
XX		ABL/71855 standard; CDNA; 292 BP.	
AC		ABL/71855;	

14-MAY-2002 (first entry)

Corn tassels-derived polynucleotide (cdps) SEQ ID NO:1229.

Corn; corn tassels-derived polynucleotide; cdps; hybrid breeding; CDps; inheritance; characteristic; growth; development; disease resistance; environmental adaptability; quality; yield; molecular marker; multigene trait; plant breeding; corn tassels; gene; ss.

Zea mays.

US2001051335-A1.

13-DEC-2001.

16-APR-1999; 99US-0294093.

21-APR-1998; 98US-082567P.

(LALG/) LALGUDI R V.
(ITOL/) ITO L Y.
(SHER/) SHERMAN B K.

Laljudi RV, Ito LY, Sherman BK:
WPI: 2002-163647/21.

Novel purified corn tassels-derived polynucleotide useful for determining altered gene expression, to recover regulatory elements and to follow inheritance of desirable characteristics through hybrid breeding programs -

Claim 1: SEQ ID 1229; 201pp: English.

The present sequence describes a purified corn tassels-derived polynucleotide sequence (cdps) comprising a nucleic acid sequence selected from those given in ABL70627 to ABL76833. The cdps sequences encode corn tassels-derived polypeptides (CDps). The cdps sequences (1) can be used for determining altered gene expression, to recover regulatory elements and to follow inheritance of desirable characteristics through hybrid breeding programs. (1) are also useful in the evaluation, and alteration of desired characteristics associated with growth and development, disease resistance, environmental adaptability, quality and yield, and as molecular markers for studying inheritance of multigene traits in a plant breeding program. (1) can be used to produce a tassels-specific profile of gene transcription, a transcript image, to clone regulatory elements for use in transformation vectors, to express a polypeptide, to identify, isolate or extend identical or related corn tassels nucleic acid sequences from DNA libraries, in nucleic acid hybridisation or amplification technologies, as query sequences to determine homology of known sequences, as probe for use in Southern or Northern hybridisation, and to identify the presence of and/or to determine the degree of similarity between two (or more) nucleic acid sequences.

Sequence 292 BP; 78 A; 56 C; 85 G; 67 T; 6 other;

Query Match 89.3%; Score 13.4; DB 24; Length 292;
Best Local Similarity 93.3%; Pred. No. 4,6e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ACCTAGCGATGCAC 15
||||| |||||||||
176 ACCTAAGCATGCAC 162

RESULT 15
AAH68618
ID AAH68618 standard; DNA: 372 BP.
AC AAH68618;
XX

DT 14-SEP-2001 (first entry)
 XX Human anti-Rh(D) chain C04 nucleotide sequence.
 DE
 XX Human; anti-Rh(D) antibody; Rh(D) binding protein; immunostimulant;
 KW red blood cell; Rh phenotype; diagnosis; therapeutic; ds.
 XX
 OS Homo sapiens.
 XX
 PN US6255455-B1.
 XX
 PD 03-JUL-2001.
 XX
 XX 29-JAN-1999; 99US-0240274.
 XX
 PF 11-OCT-1996; 96US-0028550.
 PR 10-APR-1998; 98US-0081380.
 PR 27-JUN-1997; 97US-0884045.
 XX
 XX (TYPE-) UNIV PENNSYLVANIA.
 PA
 XX
 PI Siegel DL;
 XX
 DR WPI; 2001-388931/41.
 DR P-PSDB; AAG93561.
 XX
 XX
 PT New isolated protein, preferably a human anti-Rh(D) antibody for use in
 PT diagnostics requiring a human instead of an animal antibody and in
 PT therapeutic medicine -
 XX
 PS
 XX
 XX Example 2; Column 47; 162pp; English.
 CC The present invention describes an isolated Rh(D) binding protein,
 CC preferably a human antibody, (I) having an amino acid sequence comprising
 CC one of the sequences (S) given in AAG93558 to AAG93669. (I) has
 CC immunostimulant activity, and can be used as an immune system stimulant.
 CC (I) can be used in diagnostic and therapeutic medicine. The antibodies
 CC are used in diagnostics that require human antibodies instead of animal
 CC antibodies, such as determine the Rh phenotype of human red blood cells.
 CC AAH6815 to AAH68726 represent the nucleotide sequence which encode
 CC AAG93558 to AAG93669. AAG93670 to AAG93697 represent anti-Rh(D) heavy
 CC chain CDR3 amino acid sequences which are given in the exemplification
 CC of the present invention.
 CC
 SQ Sequence 372 BP; 86 A; 92 C; 108 G; 86 T; 0 other;
 Query Match 89.3%; Score 13.4; DB 22; Length 372;
 Best Local Similarity 93.3%; Pred. No. 4.7e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 ACCTACGGCATGCAC 15
 ||||| |||||
 Db 91 ACCTATGCATGCAC 105

Search completed: June 26, 2003, 03:11:48
 Job time : 88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-25
Perfect score: 15
Sequence: 1 acctacgcattgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	4	US-08-918-148-25
2	13.4	89.3	372	4	US-09-240-274-73
3	13.4	89.3	375	4	US-09-240-274-78
4	13.4	89.3	375	4	US-09-240-274-93
5	13.4	89.3	2348	4	US-09-388-743-5
6	13.4	89.3	3735	4	US-08-975-762-43
7	13.4	89.3	3735	4	US-09-295-028-43
8	13.4	89.3	3735	4	US-09-106-582-43
9	13.4	89.3	87563	4	US-09-453-702B-57
10	12.4	82.7	24	4	US-09-673-018-6
11	12.4	82.7	191	5	PCT-US95-09114-18
12	12.4	82.7	265	4	US-08-990-823-15
13	12.4	82.7	302	3	US-08-964-268-18
14	12.4	82.7	381	4	US-09-240-274-182
15	12.4	82.7	717	4	US-09-513-783A-45
16	12.4	82.7	720	3	US-09-094-359-3
17	12.4	82.7	720	3	US-09-172-063-11
18	12.4	82.7	720	4	US-09-316-919-12
19	12.4	82.7	818	4	US-09-221-017B-900
20	12.4	82.7	850	3	US-09-062-102-2
21	12.4	82.7	850	4	US-09-364-946-2
22	12.4	82.7	972	3	US-09-172-063-29
23	12.4	82.7	1001	4	US-09-641-638-176
24	12.4	82.7	1001	4	US-09-641-638-177
25	12.4	82.7	1001	4	US-09-641-638-178
26	12.4	82.7	1095	4	US-09-085-305-5
27	12.4	82.7	1155	2	US-08-387-942C-7

28	12.4	82.7	1155	2	US-08-387-942C-18	Sequence 18, Appl
29	12.4	82.7	1176	2	US-08-387-942C-17	Sequence 17, Appl
30	12.4	82.7	1380	4	US-09-513-783A-169	Sequence 169, App
31	12.4	82.7	1419	4	US-09-516-914-4	Sequence 4, Appl
32	12.4	82.7	1515	4	US-09-080-625-6	Sequence 6, Appl
33	12.4	82.7	1515	4	US-09-695-782-6	Sequence 6, Appl
34	12.4	82.7	1521	1	US-08-726-136-20	Sequence 20, Appl
35	12.4	82.7	1521	3	US-09-103-434-20	Sequence 20, Appl
36	12.4	82.7	1521	4	US-09-687-594-20	Sequence 20, Appl
37	12.4	82.7	1586	4	US-09-673-018-1	Sequence 1, Appl
38	12.4	82.7	1650	1	US-08-459-100A-1	Sequence 1, Appl
39	12.4	82.7	1650	5	PCT-US94-09589-1	Sequence 1, Appl
40	12.4	82.7	1770	4	US-09-513-783A-1	Sequence 1, Appl
41	12.4	82.7	1929	2	US-08-818-253-1	Sequence 1, Appl
42	12.4	82.7	1929	2	US-08-818-253-5	Sequence 1, Appl
43	12.4	82.7	1929	4	US-08-818-252-1	Sequence 1, Appl
44	12.4	82.7	1929	4	US-08-818-252-5	Sequence 1, Appl
45	12.4	82.7	1959	2	US-08-818-253-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-08-918-148-25
: Sequence 25, Application US/08918148A
: Patent No. 6342220
: GENERAL INFORMATION:
: APPLICANT: Adams, Camellia
: APPLICANT: W.
: APPLICANT: Carter, Paul J.
: APPLICANT: Fendly, Brian M.
: APPLICANT: Gurney, Austin L.
: TITLE OF INVENTION: Agonist Antibodies
: FILE REFERENCE: P0979
: CURRENT APPLICATION NUMBER: US/08/918,148A
: CURRENT FILING DATE: 1997-08-25
: NUMBER OF SEQ ID NOS: 79
: SEQ ID NO 25
: LENGTH: 15
: TYPE: DNA
: ORGANISM: artificial
: FEATURE:
: NAME/KEY: 10D10scfv, 12B5scfv VH CDR1
: LOCATION: 1-15
: OTHER INFORMATION:
US-08-918-148-25

Query Match      100.0%; Score 15; DB 4; Length 15;
Best local Similarity 100.0%; Pred. No. 10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCTACGCATGCAC 15
DB      1 ACCTACGCATGCAC 15

RESULT 2
US-09-240-274-73
: Sequence 73, Application US/09240274
: Patent No. 625455
: GENERAL INFORMATION:
: APPLICANT: Siegel, Donald L.
: TITLE OF INVENTION: RH(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
: TITLE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
: FILE REFERENCE: 09596-42U2
: CURRENT APPLICATION NUMBER: US/09/240,274
: CURRENT FILING DATE: 1999-01-29
: EARLIER APPLICATION NUMBER: 60/081,380
: EARLIER FILING DATE: 1998-04-10
: EARLIER APPLICATION NUMBER: 60/028,550
: EARLIER FILING DATE: 1996-10-11
: NUMBER OF SEQ ID NOS: 224
```

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 73
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain C04
US-09-240-274-73

Query Match          89.3%; Score 13.4; DB 4; Length 372;
Best Local Similarity 93.3%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCAC 15
Db      91 ACCTATGGCATGCAC 105

RESULT 3
US-09-240-274-78
; Sequence 78, Application US/09240274
; Patent No. 6255455
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/240,274
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/081,380
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/028,550
; EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 78
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain D03
US-09-240-274-78

Query Match          89.3%; Score 13.4; DB 4; Length 375;
Best Local Similarity 93.3%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCAC 15
Db      91 ACCTATGGCATGCAC 105

RESULT 4
US-09-240-274-93
; Sequence 93, Application US/09240274
; Patent No. 6255455
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/240,274
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/081,380
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/028,550
; EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 93
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain D20
US-09-240-274-93

Query Match          89.3%; Score 13.4; DB 4; Length 375;
Best Local Similarity 93.3%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCAC 15
Db      91 ACCTATGGCATGCAC 105

RESULT 5
US-09-388-743-5/C
; Sequence 5, Application US/09388743
; Patent No. 6423886
; GENERAL INFORMATION:
; APPLICANT: Singletary, George
; APPLICANT: Zhou, Ian
; TITLE OF INVENTION: No. 6423886el Starch Synthase Polynucleotides and Their
; FILE REFERENCE: 1144
; CURRENT APPLICATION NUMBER: US/09/388,743
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2348
; TYPE: DNA
; ORGANISM: Curcuma zedoaria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (36)...(2105)
US-09-388-743-5

Query Match          89.3%; Score 13.4; DB 4; Length 2348;
Best Local Similarity 93.3%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCAC 15
Db      1868 ACCTACTGCATGCAC 1854

RESULT 6
US-08-975-762-43
; Sequence 43, Application US/08975762
; Patent No. 6207169
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,762
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
```

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 3735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-975-762-43

Query Match 89.3%; Score 13.4; DB 4; Length 3735;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGCATGCAC 15
||||| |||||||
Db 1390 ACCTACGCATGCAC 1404

RESULT 7
US-09-295-028-43
Sequence 43, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 43
LENGTH: 3735
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-43

Query Match 89.3%; Score 13.4; DB 4; Length 3735;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGCATGCAC 15
||||| |||||||
Db 1390 ACCTACGCATGCAC 1404

RESULT 8
US-09-106-582-43
Sequence 43, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: 582
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 3735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-106-582-43

Query Match 89.3%; Score 13.4; DB 4; Length 3735;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGCATGCAC 15
||||| |||||||
Db 1390 ACCTACGCATGCAC 1404

RESULT 9
US-09-453-702B-57/c
Sequence 57, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 87563

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-09-453-702B-57

Query Match 89.3%; Score 13.4; DB 4; Length 87563;
Best Local Similarity 93.3%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACTACGGCATGCAC 15
Db 41493 ACTACGGCATGCAC 41479

RESULT 10
US-09-673-018-6/c
Sequence 6, Application US/09673018
Patent No. 6461842
GENERAL INFORMATION:
APPLICANT: MATSUDA, Hideyuki
APPLICANT: KANAMUKAI, Makoto
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NISHI, Kenichi
APPLICANT: HASEGAWA, Junzo
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: METHOD FOR PREPARING COENZYME Q10
FILE REFERENCE: 2000-1379A/LC/00177
CURRENT APPLICATION NUMBER: US/09/673.018
CURRENT FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: JP 11/32657
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 24
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA Primers
US-09-673-018-6

Query Match 82.7%; Score 12.4; DB 4; Length 24;
Best Local Similarity 92.9%; Pred. No. 2.8e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CCTACGGCATGCAC 15
Db 19 CCTACGGCATGCAC 6

RESULT 11
PCT-US95-09114-18
Sequence 18, Application PC/TUS9509114
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Antibodies that Bind a Conformationally
TITLE OF INVENTION: Altered CD4 Molecule Induced Upon Binding of
TITLE OF INVENTION: Human Immunodeficiency Virus
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
City: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09114
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,080
FILING DATE: 19-Jul-94
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305,903
FILING DATE: 13-Sep-94
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: B1Z-013CPC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US95-09114-18

Query Match 82.7%; Score 12.4; DB 5; Length 191;
Best Local Similarity 92.9%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACTACGGCATGCA 14
Db 91 AGCTACGGCATGCA 104

RESULT 12
US-08-990-823-15/c
Sequence 15, Application US/08990823D
Patent No. 6228371
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990,823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 15
LENGTH: 265
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-15

Query Match 82.7%; Score 12.4; DB 4; Length 265;
Best Local Similarity 92.9%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CCTACGGCATGCAC 15
Db 33 CGTACGGCATGCAC 20

RESULT 13
US-08-964-268-18
Sequence 18, Application US/08964268
Patent No. 6114503
GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: RUBEN, STEVEN M
APPLICANT: SANCAR, AZIZ
APPLICANT: HSU, SHIO-WEN D
APPLICANT: KAZANTSEV, ALEKSEY G
TITLE OF INVENTION: HUMAN BLUE-LIGHT PHOTORECEPTOR hCRY2
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,268
FILING DATE: HEREWITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030,189
FILING DATE: 04-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0750001/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-964-268-18

Query Match 82.7%: Score 12.4; DB 3; Length 302;
Best Local Similarity 86.7%: Pred. No. 3.1e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACCTACGCGATGCAC 15
|||||
Db 126 ACCTACGCGGTGCMC 140

RESULT 14
US-09-240-274-182
Sequence 182, Application US/09240274
Patent No. 6255455
GENERAL INFORMATION:
APPLICANT: Siegel, Donald L.
TITLE OF INVENTION: RH(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
FILE REFERENCE: 09596-4202
CURRENT APPLICATION NUMBER: US/09/240,274
CURRENT FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: 60/081,380
EARLIER FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/028,550
EARLIER FILING DATE: 1996-10-11
NUMBER OF SEQ ID NOS: 224
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 182

LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: anti-Rh(D) antibody clone SH10
US-09-240-274-182

Query Match 82.7%: Score 12.4; DB 4; Length 381;
Best Local Similarity 92.9%: Pred. No. 3.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTACGCGATGCAC 15
|||||
Db 332 CCTACGCGATGCAC 345

RESULT 15
US-09-513-783A-45
Sequence 45, Application US/09513783A
Patent No. 6416959
GENERAL INFORMATION:
APPLICANT: Giuliano, Kenneth A.
APPLICANT: Kapur, Ravi
TITLE OF INVENTION: A System for Cell Based Screening
FILE REFERENCE: 97-022-L1
CURRENT APPLICATION NUMBER: US/09/513,783A
CURRENT FILING DATE: 2000-02-25
NUMBER OF SEQ ID NOS: 180
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 45
LENGTH: 717
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(717)
OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-45

Query Match 82.7%: Score 12.4; DB 4; Length 717;
Best Local Similarity 92.9%: Pred. No. 3.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGCGATGCA 14
|||||
Db 196 ACCTACGCGGTGCA 209

Search completed: June 26, 2003, 04:52:26
Job time: 30.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-25

Perfect score: 15
Sequence: 1 aactcgcgcacgac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	93.3	1575	9	US-10-027-806-9
2	14	93.3	1575	9	US-10-034-623-9
3	14	93.3	1575	9	US-10-027-801-9
4	14	93.3	42432	9	US-10-027-806-2
5	14	93.3	42432	9	US-10-034-623-2
6	14	93.3	42432	9	US-10-027-801-2
7	13.4	89.3	292	10	US-09-294-093B-1229
8	13.4	89.3	372	9	US-09-848-798-73
9	13.4	89.3	375	9	US-09-848-798-78
10	13.4	89.3	375	9	US-09-848-798-93
11	13.4	89.3	375	9	US-10-172-317-1
12	13.4	89.3	714	9	US-09-822-846-612
13	13.4	89.3	933	10	US-09-838-561-6
14	13.4	89.3	933	10	US-09-816-760-6
15	13.4	89.3	1157	10	US-09-820-893-31
16	13.4	89.3	1379	10	US-09-838-561-4
17	13.4	89.3	1379	10	US-09-816-760-4
18	13.4	89.3	1743	9	US-09-738-626-1827
19	13.4	89.3	3735	10	US-09-159-469-43

20	13.4	89.3	3735	10	US-09-798-042-43	Sequence 43, Appl
21	13.4	89.3	3735	10	US-09-798-042-88	Sequence 88, Appl
22	13.4	89.3	3735	10	US-09-798-042-96	Sequence 96, Appl
23	13.4	89.3	87563	9	US-10-114-170-57	Sequence 57, Appl
24	13.4	89.3	3309400	9	US-09-738-626-1	Sequence 1, Appl
25	13	86.7	477	9	US-09-738-626-1449	Sequence 1449, Ap
26	13	86.7	498	9	US-09-918-995-11630	Sequence 11630, A
27	13	86.7	1311	10	US-09-925-297-354	Sequence 354, App
28	12.4	82.7	53	10	US-09-962-628B-19	Sequence 19, Appl
29	12.4	82.7	53	10	US-09-962-628B-20	Sequence 20, Appl
30	12.4	82.7	53	10	US-09-962-628B-22	Sequence 22, Appl
31	12.4	82.7	144	10	US-09-294-093B-3458	Sequence 3458, Ap
32	12.4	82.7	157	10	US-09-864-761-30169	Sequence 30169, A
33	12.4	82.7	265	9	US-09-996-634-15	Sequence 15, Appl
34	12.4	82.7	265	9	US-09-997-181-15	Sequence 15, Appl
35	12.4	82.7	265	9	US-09-997-182-15	Sequence 15, Appl
36	12.4	82.7	293	10	US-09-864-761-24230	Sequence 24230, A
37	12.4	82.7	378	9	US-09-835-976B-69	Sequence 69, Appl
38	12.4	82.7	381	9	US-09-848-798-182	Sequence 182, App
39	12.4	82.7	388	9	US-09-954-531-724	Sequence 724, App
40	12.4	82.7	462	10	US-09-770-444-404	Sequence 404, App
41	12.4	82.7	536	10	US-09-974-300-955	Sequence 955, App
42	12.4	82.7	558	10	US-09-864-761-7525	Sequence 7525, Ap
43	12.4	82.7	575	10	US-09-864-761-13641	Sequence 13641, A
44	12.4	82.7	717	9	US-09-989-025A-1	Sequence 1, Appl
45	12.4	82.7	717	9	US-10-100-957A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-10-027-806-9
; Sequence 9, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOMP 002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; US-10-027-806-9

Query Match 93.3%; Score 14; DB 9; Length 1575;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTACGCGATGCAC 15
Db 332 CCTACGCGATGCAC 345
US-10-034-623-9
; Sequence 9, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa

```

; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
US-10-034-623-9

```

```

Query Match          93.3%; Score 14; DB 9; Length 1575;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      2 CCTACGGCATGCAC 15
        |||
Db      332 CCTACGGCATGCAC 345

```

```

RESULT 3
US-10-027-801-9
; Sequence 9, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
US-10-027-801-9

```

```

Query Match          93.3%; Score 14; DB 9; Length 1575;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      2 CCTACGGCATGCAC 15
        |||
Db      332 CCTACGGCATGCAC 345

```

```

RESULT 4
US-10-027-806-2
; Sequence 2, Application US/10027806
; Patent No. US20020160476A1.
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806

```

```

; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
US-10-027-806-2

```

```

Query Match          93.3%; Score 14; DB 9; Length 42432;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      2 CCTACGGCATGCAC 15
        |||
Db      13377 CCTACGGCATGCAC 13390

```

```

RESULT 5
US-10-034-623-2
; Sequence 2, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS

```

```

; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
US-10-034-623-2

Query Match          93.3%; Score 14; DB 9; Length 42432;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 CCTACGGCATGCAC 15
DB      13377 CCTACGGCATGCAC 13390

RESULT 6
US-10-027-801-2
; Sequence 2, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCOEP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 42432
TYPE: DNA
ORGANISM: Cenarchaeum symbiosum
FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (11046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
```

```

; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
US-10-027-801-2

Query Match          93.3%; Score 14; DB 9; Length 42432;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 CCTACGGCATGCAC 15
DB      13377 CCTACGGCATGCAC 13390

RESULT 7
US-09-294-093B-1229/C
; Sequence 1229, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
SEQ ID NO 1229
LENGTH: 292
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
; NAME/KEY: misc. feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343977H1
; LOCATION: 52, 62, 68, 83, 90, 145
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-1229

Query Match          89.3%; Score 13.4; DB 10; Length 292;
Best Local Similarity 93.3%; Pred. No. 26+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ACCTACGGCATGCAC 15
DB      176 ACCTACGGCATGCAC 162

RESULT 8
US-09-848-798-73
; Sequence 73, Application US/09848798
; Publication No. US20030040605A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rn(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/848,798
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/240,274
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/028,550
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 73
LENGTH: 372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: anti-Rn(D) chain C04
```

US-09-848-798-73

Query Match 89.3%; Score 13.4; DB 9; Length 372;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTAGGCGATGCAC 15
Db 91 ACCTATGGCATGCAC 105

US-09-848-798-78

Sequence 78, Application US/09848798
Publication No. US20030040605A1
GENERAL INFORMATION:

APPLICANT: Siegel, Donald L.
TITLE OF INVENTION: Rn(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
FILE REFERENCE: 09596-4202
CURRENT FILING DATE: 2001-05-04
PRIOR APPLICATION NUMBER: US/09/848,798
PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/240,274
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1999-01-29
PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 60/028,550
NUMBER OF SEQ ID NOS: 224
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 78
LENGTH: 375
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: anti-Rh(D) chain D03
US-09-848-798-78

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTAGGCGATGCAC 15
Db 91 ACCTATGGCATGCAC 105

RESULT 10

US-09-848-798-93
Sequence 93, Application US/09848798
Publication No. US20030040605A1
GENERAL INFORMATION:
APPLICANT: Siegel, Donald L.
TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
FILE REFERENCE: 09596-4202
CURRENT FILING DATE: 2001-05-04
PRIOR APPLICATION NUMBER: US/09/848,798
PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/240,274
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1999-01-29
PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 60/028,550
NUMBER OF SEQ ID NOS: 224
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 93
LENGTH: 375
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: anti-Rh(D) chain D20
US-09-848-798-93

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTAGGCGATGCAC 15
Db 91 ACCTATGGCATGCAC 105

RESULT 11

US-10-172-317-1
Sequence 1, Application US/10172317
Publication No. US20030091561A1
GENERAL INFORMATION:
APPLICANT: van de Winkel, Jan G.J.
APPLICANT: van Dijk, Marcus A.
APPLICANT: Halk, Edward
APPLICANT: Gerlitsen, Arnout F.
TITLE OF INVENTION: HUMAN MONOCLONAL ANTIBODIES TO
FILE REFERENCE: GMI-020
CURRENT FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US/10/172,317
PRIOR FILING DATE: 2001-06-13
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 375
TYPE: DNA
ORGANISM: Homo sapiens
US-10-172-317-1

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTAGGCGATGCAC 15
Db 91 ACCTATGGCATGCAC 105

RESULT 12

US-09-822-846-612/C
Sequence 612, Application US/09822846
Publication No. US20030027139A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John W.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Yikl
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakara
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES' ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6400
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: US/09/822,846
PRIOR APPLICATION NUMBER: 60/195,605
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 629
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 612
LENGTH: 714

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-612

Query Match      89.3%; Score 13.4; DB 9; Length 714;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ACCTACGGCATGCAC 15
DB      219 ACCTACGGCATGCAC 205

RESULT 13
US-09-838-561-6
; Sequence 6, Application US/09838561
; Patent No. US2002042371A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Cook, William James
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; APPLICANT: Gimeno, Ruth
; TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL DEHYDROGENASE
; FILE REFERENCE: MNI-134CP2
; CURRENT APPLICATION NUMBER: US/09/838,561
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 09/816,760
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 09/634,955
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 60/192,002
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(933)
US-09-838-561-6

Query Match      89.3%; Score 13.4; DB 10; Length 933;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ACCTACGGCATGCAC 15
DB      745 ACCTACGGCATGCAC 759

RESULT 14
US-09-816-760-6
; Sequence 6, Application US/09816760
; Patent No. US20020052032A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Cook, William James
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL HUMAN DEHYDROGENASE
; FILE REFERENCE: MNI-134CP
; CURRENT APPLICATION NUMBER: US/09/816,760
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 09/634,955
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 60/192,002
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 13
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(933)
US-09-816-760-6

Query Match      89.3%; Score 13.4; DB 10; Length 933;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ACCTACGGCATGCAC 15
DB      745 ACCTACGGCATGCAC 759

RESULT 15
US-09-820-893-31
; Sequence 31, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1157
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-820-893-31

Query Match      89.3%; Score 13.4; DB 10; Length 1157;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ACCTACGGCATGCAC 15
DB      848 ACCTACGGCATGCAC 862
```

Search completed: June 26, 2003, 04:57:12
Job time : 65.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: us-09-138-091-25

Perfect score: 15

Sequence: 1 acctacgcatgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15	100.0	354	14	R12983 yf70e11.r1
C 2	15	100.0	549	12	BF326887 QV3-BN004
C 3	15	100.0	663	12	BG539077 602568449
C 4	15	100.0	706	14	BM716521 UI-E-E30-
C 5	15	100.0	923	12	BE882505 601507246
C 6	15	100.0	961	14	BO719668 AGENCOURT

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
R12983	354 bp	mrna	linear	EST 12-APR-1995								
IMAGE:27463 5												
R12983	GI:766059											
Homo sapiens												
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;												
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.												
1 (bases 1 to 354)												
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman												
, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,												
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston												
, R., Williamson, A., Woldmann, P. and Wilson, R.												
The WashU-Merck EST Project												
Unpublished (1995)												
Contact: Wilton RK												
Washington University School of Medicine												
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108												
Tel: 314 286 1800												
Fax: 314 286 1810												
Email: est@wustl.edu												
Insert Size: 2106												
High quality sequence stops: 175 Source: IMAGE Consortium, LNL												
This clone is available royalty-free through LNL												
IMAGE Consortium (info@image.lnl.gov) for further information.												

ALIGNMENTS

7	14	93.3	268	14	T82729
C 8	14	93.3	325	9	AT764493
C 9	14	93.3	369	13	B1183951
C 10	14	93.3	486	14	BM728351
C 11	14	93.3	531	14	BO040153
C 12	14	93.3	553	17	AO657745
C 13	14	93.3	569	10	AV386289
C 14	14	93.3	591	17	AO943516
C 15	14	93.3	591	17	AZ103592
C 16	14	93.3	591	17	AO652602
C 17	14	93.3	602	17	AO652520
C 18	14	93.3	608	10	AV385351
C 19	14	93.3	616	17	AO660685
C 20	14	93.3	674	10	AV385001
C 21	14	93.3	696	17	BH545455
C 22	14	93.3	761	17	BH696405
C 23	14	93.3	877	12	BF028074
C 24	14	93.3	1096	12	BG474704
C 25	13.4	89.3	105	12	BE860369
C 26	13.4	89.3	105	17	BH865555
C 27	13.4	89.3	137	17	AO646650
C 28	13.4	89.3	162	10	AW984704
C 29	13.4	89.3	200	9	AU076815
C 30	13.4	89.3	203	14	D30996
C 31	13.4	89.3	209	10	BE365107
C 32	13.4	89.3	213	10	BE157960
C 33	13.4	89.3	216	12	BF088542
C 34	13.4	89.3	231	17	AZ311664
C 35	13.4	89.3	248	14	BM784367
C 36	13.4	89.3	248	14	H00321
C 37	13.4	89.3	264	10	AV328192
C 38	13.4	89.3	274	14	F35823
C 39	13.4	89.3	288	13	BM193261
C 40	13.4	89.3	289	10	BB489547
C 41	13.4	89.3	291	10	AM461362
C 42	13.4	89.3	297	14	T27868
C 43	13.4	89.3	307	9	A1470361
C 44	13.4	89.3	307	17	BH239762
C 45	13.4	89.3	308	17	FR0003712

Insert Length: 2106	Std Error: 0.00
Seq primer: M13RPl	
High quality sequence stop: 175.	

BASE COUNT	61 a	83 c	107 g	99 t	4 others
ORIGIN					

RESULT 2				
BF326887/C				
LOCUS	BF326887	549 bp	mRNA	linear
DEFINITION	OVA-BN0046-060700-256-d10	BN0046	Homo sapiens	CDNA, mRNA sequence.
ACCESSION	BF326887			
VERSION	BF326887.1	GI:11297635		
KEYWORDS	EST.			
SOURCE	human.			

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BN0046"
/dev_stage="Adult"
/name="Organ: breast_normal; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORSTES PCR (U.S. letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      113 a      133 c      156 g      147 t
ORIGIN

```

RESULT		663 bp	mRNA	linear	EST 03-APR-2001
BG539077/c					
LOCUS	BG539077				
DEFINITION	BG539077 602586449P1 NIH_MGC_77	Homo sapiens CDNA clone IMAGE:4692781 5'			
ACCESSION	BG539077				
VERSION	BG539077.1	GI:1351310			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				

FEATURES	Location/Qualifiers
source	1. .549

BASE COUNT	146 a	131 c	178 g	208 t
ORIGIN				
Query Match		100.0%;	Score 15;	DB 12;
Best local Similarity		100.0%;	Pred. No. 1.6e+03;	Length 663;
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 ACCTACGCATGCAC 15
 |||||||
 Db 92 ACCTACGCATGCAC 78

RESULT 4
 BM716521/c 706 bp mRNA linear EST 28-FEB-2002
 LOCUS BM716521
 DEFINITION U1-E-EJ0-ah1-i-11-0-U1-r2 U1-E-EJ0 Homo sapiens cDNA clone
 U1-E-EJ0-ah1-i-11-0-U1 5', mRNA sequence.

ACCESSION BM716521
 VERSION BM716521
 KEYWORDS GI:19029779
 SOURCE EST.
 ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS 1 (bases 1 to 706)
 TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 JOURNAL Normalization and subtraction: two approaches to facilitate gene
 9704477
 MEDLINE
 COMMENT

Genome Res. 6 (9), 791-806 (1996)
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hegeman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers

1..706
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="U1-E-EJ0-ah1-i-11-0-U1"
 /clone_1b="U1-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 U1-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AATAATCAAGA
 ; lens, CGATTAGCGA; eye anterior segment, AATGCCCAT;
 optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and
 Macular, GTCC; RPE and Choroid, ACCCTA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT 183 a 130 c 164 g 228 t 1 others

Query Match 100.0%; Score 15; DB 14; Length 706;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGCATGCAC 15
 |||||||
 Db 28 ACCTACGCATGCAC 14

RESULT 5
 BE882505/c 923 bp mRNA linear EST 20-OCT-2000
 LOCUS BE882505
 DEFINITION 601507246F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3908891 5',
 mRNA sequence.

ACCESSION BE882505
 VERSION BE882505
 KEYWORDS GI:10331281
 SOURCE EST.
 ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS 1 (bases 1 to 923)
 TITLE NIH-MGC http://mhc.nci.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 9704477
 MEDLINE
 COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM9721 row: 1 column: 12
 High quality sequence stop: 608.

FEATURES
 source Location/Qualifiers

1..923
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3908891"
 /clone_1b="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 Kb.
 170 a 228 c 255 g 270 t

Query Match 100.0%; Score 15; DB 12; Length 923;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGCATGCAC 15
 |||||||
 Db 478 ACCTACGCATGCAC 464

RESULT 6
 B0719668/c 961 bp mRNA linear EST 16-JUL-2002
 LOCUS B0719668
 DEFINITION AGENCOUNT 8304828 lupskl.sympathetic_trunk Homo sapiens cDNA clone
 IMAGE:6193699 5', mRNA sequence.

ACCESSION B0719668
 VERSION B0719668
 KEYWORDS GI:21858565
 SOURCE EST.
 ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS 1 (bases 1 to 961)
 TITLE NIH-MGC http://mhc.nci.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 9704477
 MEDLINE
 COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

67 a 68 c 69 g 64 t

/note=Vector: pYES2; Site_1: XhoI; Site_2: EcoRI; cDNA
was synthesized with ZAP kit (Stratagene) using poly(A)+
RNA isolated from *Aedes albopictus* niger grown in both complete
and minimal media. Synthesis was primed with oligo(dT)
primer/XhoI-linker. EcoRI adaptors were later ligated to
polished ends. EcoRI-XhoI-digested cDNA was ligated with
EcoRI-XhoI-digested pYES2 (Invitrogen Corp). This vector
permits expression of our library in yeast.

Query Match	93.3%	Score 14	DB 14	Length 268
Best Local Similarity	100.0%	Pred. No. 3	6e+03	
Matches	14	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
Qy	2	CCTACGGCATGCAC	15	
Db	45	CCTACGGCATGCAC	58	
RESULT 8	AT764493/c			
LOCUS	AT764493	325 bp	mRNA	linear
DEFINITION	UI-R-Y0-abk-a-10-0-UI s1 UI-R-Y0 Rattus norvegicus cDNA clone			EST 25-JUN-1999
ACCESSION	AT764493			
VERSION	AT764493.1			
KEYWORDS	EST.			
SOURCE	Norway rat.			

REFERENCE 1 (bases 1 to 325)
AUTHORS Ronaldo, M. F., Lennon, G. and Soares, M. B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
MEDLINE Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
Contact: Soares, MB

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dt track not found. Not a site shown in beginning of sequence
is likely internal to the message. CDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 95-157, >(CGG
)n/simple-repeat
Seq primer: M13 Forward
POLYA-NO.

```

/organism="Aspergillus niger"
/db_xref="taxon:5061"
/clone_lib="Aspergillus niger, pYES2 (XhoI-EcORI)"
/lab_host="E. coli"

```

```

Ch          93.38;   Score 14;   DB 14;   Length 268;
1 Similarity 100.0%;   Pred. No. 3,6e+03;
14;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

2 CCTACGCGCATGCAC 15
   |||||||||||
45 CCTACGCGCATGCAC 58

A1764493          325 bp      mRNA      linear      EST 25-JUN-1999
UI-R-Y0-abk-a-10-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
UI-R-Y0-abk-a-10-0-UI 3, mRNA sequence.
A1764493
A1764493.1 GI:5210386
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 325)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 95-157, >(CGG
)n/simple-repeat
Seq primer: M13 Forward
POLYA-No.

Location/Qualifiers
1. 325
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-abk-a-10-0-UI"
/clone_11b="UI-R-Y0"
/dev_stage="adult"
/lab_host="PH10B (Life Technologies)"
/note="vector: pUT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted

```

library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 37 a 118 c 126 g 43 t 1 others
 ORIGIN TAG_LIB=UI-R-Y0
 TAG_TISSUE=Eye
 TAG_SEQ=CATTG"

Query Match 93.3%; Score 14; DB 9; Length 325;
 Best Local Similarity 100.0%; Pred. No. 3.9e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
 Db 174 CCTACGGCATGCAC 161
 |||

RESULT 9 369 bp mRNA linear EST 10-JUL-2001
 B1183951 UNL-P-FN-cf-e-03-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 LOCUS UNL-P-FN-cf-e-03-0-UNL 3', mRNA sequence.
 DEFINITION B1183951
 ACCESSION B1183951
 VERSION B1183951.1 GI:14658360
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 369)
 Caetano, A.R., Johnson, R.K. and Pomp, D.
 Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
 Unpublished (2001)
 JOURNAL Contact: Pomp, D
 COMMENT Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag served to verify it as a clone from the normalized porcine ovarian follicles library. The following repetitive elements were found in this cDNA sequence: 258-349, >1120871emb1X64127.11SSPRE.S.scrofa DNA for SINE sequence SSPRE
 Seq primer: M13 -29
 POLY-A:yes.

FEATURES
 source Location/Qualifiers
 1..369
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-cf-e-03-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT733-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; The UNL-P-FN

library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996.

BASE COUNT 86 a 100 c 99 g 84 t
 ORIGIN TAG_LIB=UNL-P-FN
 TAG_TISSUE=porcine ovarian follicles
 TAG_SEQ=CACACT"

Query Match 93.3%; Score 14; DB 13; Length 369;
 Best Local Similarity 100.0%; Pred. No. 4.1e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCA 14
 Db 264 ACCTACGGCATGCA 277
 |||

RESULT 10 486 bp mRNA linear EST 01-MAR-2002
 BM728351 UI-E-E01-ai-v-i-15-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone
 LOCUS UI-E-E01-ai-v-i-15-0-UI 5', mRNA sequence.
 DEFINITION BM728351
 ACCESSION BM728351
 VERSION BM728351.1 GI:19049684
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 486)
 Bonaldi, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@iuiweeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hagaman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers
 1..486
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-E01-ai-v-i-15-0-UI"
 /clone_lib="UI-E-E01"
 /tissue_type="fetal eye"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT733-Pac (Pharmacia) with a modified polylinker. Site_1: EcoR I; Site_2: Not I; UI-E-E01 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6: 791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I

BASE COUNT	ORIGIN
110 a	110 c
136 g	127 t
	3 others

OY 1 AACCTACGGCATGCAC 15
| | | | | | | | | |
Db 444 ACNTACGGCATGCAC 430

[illegible]

ORGANISM

AUTHORS

COMMENT

Email: est@watson.wustl.edu

Libraries were constructed by Dr Stavros Bashlindas as part of the Physcometrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA). DNA sequencing by: Washington University Genome Sequencing Center. For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk). High quality sequence stop: 421.

source

```

1. 531
/organism="Physcomitrella patens"
/db_xref="taxon:3218"
/clone="PEP-SOURCE_ID:PPS40315"
/clone_11b="Moss EST library PPS"
/dev_stage="prothematia, 7day old untreated"
/lab_host="E.coli DH10b"
/notes="vector: pbluescript SK-; Site_1: XhoI; Site_2:
EcoRI; library constructed by Stavros Bashlaidis and
re-arranged by A. Cumming & Honglin Rong. Construction of
the cDNA library was carried out using Stratagene's 'UniZAP-
cDNA synthesis kit' to ligate cDNA directionally in
UniZAP XR vector arms. The vector is designed containing
the pbluescript sequence as well as the lambda DNA and
cDNA is cloned in the EcoRI and XhoI sites in the
pbluescript sequence. The vector was then packaged using
cold glycopackaging extracts, propagated in XL-Ibue MRF

```

BASE COUNT	104 a	135 c	145 g	147 t
------------	-------	-------	-------	-------

OY	1	ACCTACGGCATGCA	14
Db	350	ACCTACGGCATGCA	363

A0657745

ADDENDUM

KEYWORDS

TOPEKA

Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Shared DNA end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.
Seq primer: M13-Forward
Class: Shotgun.

FEATURES

Source

/organism="Trypanosoma brucei"
 /strain="TRE927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-2796"
 /clone_1fb="Sheared DNA"
 /note="Vector: pUC18; Site_1: SmaI. Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TRE927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaudin and B. Barrell, Oxford University
 Press, 1999).".

BASE COUNT 182 a 149 c 107 g 115 t
ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 553;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ACCTACGCATGCA 14
| | | | | | | | | | | | | | | |
Db 325 ACCTACGCATGCA 338

RESULT 13
AV386289 569 bp mRNA linear EST 27-OCT-1999
LOCUS AV386289 Halocynthia roretzi Fertilized egg Halocynthia roretzi
DEFINITION CDNA clone 009B23_3', mRNA sequence.
ACCESSION AV386289
VERSION AV386289.1 GI:6131346
KEYWORDS EST.
SOURCE Halocynthia roretzi.
ORGANISM Halocynthia roretzi.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE 1 (bases 1 to 569)
AUTHORS Makabe, K.W.
TITLE Halocynthia roretzi EST
JOURNAL Unpublished (1999)
CONTACT: Kazuhiro W. Makabe
Department of Zoology, Graduate School of Science
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: kwmakabe@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source Location/Qualifiers
1..569
/organism="Halocynthia roretzi"
/db_xref="taxon:7729"
/clone="009B23_3"
/dev_stage="Fertilized egg"
/note="Organ: embryo"

BASE COUNT 157 a 126 c 124 g 162 t
ORIGIN

Query Match 93.3%; Score 14; DB 10; Length 569;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 CCTACGCATGCAC 15
| | | | | | | | | | | | | | | |
Db 160 CCTACGCATGCAC 173

RESULT 14
AO943516 591 bp DNA linear GSS 27-JAN-2000
LOCUS AO943516
DEFINITION Sheared DNA-53M15.TR Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AO943516
VERSION AO943516.1 GI:6766781
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 591)
AUTHORS El-Sayed, N., Zhao, S., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library

JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-53M15.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tldb/tldb/>.
Seq primer: M13-Reverse
Class: Shotgun.

FEATURES
source Location/Qualifiers
1..591
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-53M15"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaundin and B. Barrell, Oxford University
Press, 1999)."

BASE COUNT 154 a 158 c 142 g 136 t 1 others
ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 591;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ACCTACGCATGCA 14
| | | | | | | | | | | | | | | |
Db 82 ACCTACGCATGCA 95

RESULT 15
AZ103592 591 bp DNA linear GSS 09-MAY-2000
LOCUS AZ103592
DEFINITION RPCI-23-21H20.TJ RPCI-23 Mus musculus genomic clone RPCI-23-21H20,
DNA sequence.
ACCESSION AZ103592
VERSION AZ103592.1 GI:7756650
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 591)
AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akınret,
B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shao Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Reseach Genetics (http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html)
 Plate: 21 row: H column: 20

Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..591
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-21H20"
 /clone_1b="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 " "

BASE COUNT 127 a 133 c 133 g 198 t

ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 591;

Best Local Similarity 100.0%; Pred. No. 4.9e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTACGGCATGCAC 15

Db 242 CCTACGGCATGCAC 229

Search completed: June 26, 2003, 04:50:14
 Job time : 754.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds

(Without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-36

Perfect score: 15

Sequence: 1 agccataactgaac 15

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rtd:*
- 36: em_hlg_mam:*
- 37: em_hlg_vrt:*
- 38: em_sy:*
- 39: em_hlgo_hum:*
- 40: em_hlgo_mus:*
- 41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	6	AR183492
2	15	100.0	43523	9	AC024250 Homo sapi
3	15	100.0	50210	9	AC079261 Homo sapi
4	15	100.0	59465	8	AY007367 Lycopersi
5	15	100.0	91240	2	AC106140 Rattus no
6	15	100.0	100685	8	AC023754 Arabidops
7	15	100.0	101282	2	AC121474 Rattus no
8	15	100.0	101290	2	AC009240 Homo sapi
9	15	100.0	108610	2	AC105725 Rattus no
10	15	100.0	121212	9	HS135L22
11	15	100.0	127400	2	AC131121 Mus muscu
12	15	100.0	129547	8	AC006434 Genomic s
13	15	100.0	145520	2	AC016036 Homo sapi
14	15	100.0	154084	9	AC022311 Homo sapi
15	15	100.0	158861	2	AC130009 Rattus no
16	15	100.0	165733	2	AC099146 Rattus no
17	15	100.0	170717	2	AC128590 Rattus no
18	15	100.0	175518	9	AC018846 Homo sapi
19	15	100.0	180613	2	AC115064 Mus muscu
20	15	100.0	182604	2	AC127446 Rattus no
21	15	100.0	182891	2	AC101658 Mus muscu
22	15	100.0	205202	2	AC108796 Mus muscu
23	15	100.0	207473	2	AC117630 Mus muscu
24	15	100.0	221372	2	AC124474 Mus muscu
25	15	100.0	227010	2	AC113490 Mus muscu
26	14	93.3	127	9	HUMRB
27	14	93.3	258	9	HSNJ9519 Homo sapi
28	14	93.3	373	8	AY038113 Montanoa
29	14	93.3	415	8	AY038114 Montanoa
30	14	93.3	415	8	AY038115 Montanoa
31	14	93.3	652	11	PM7H5B
32	14	93.3	735	11	PM3B9G
33	14	93.3	2570	8	AF422179 Arabidops
34	14	93.3	2898	1	L1095056 Leprosifira
35	14	93.3	3456	3	AY051923 Drosophill
36	14	93.3	9908	1	AE013309 Methanosa
37	14	93.3	23819	2	AC018173 Drosophill
38	14	93.3	36138	1	MJ115187 U15187 Mycobacteri
39	14	93.3	38109	1	MLCB5
40	14	93.3	42936	3	U53333 Mycobacteri
41	14	93.3	60328	9	AL500526 Caenorhabdi
42	14	93.3	63178	3	AC004333 Human DNA
43	14	93.3	64631	2	AC004333 Drosophill
44	14	93.3	64767	9	AC118478 Mus muscu
45	14	93.3	71320	2	U69730 Homo sapien
					AC095414 Rattus no

ALIGNMENTS

RESULT 1
AR183492 LOCUS AR183492 15 bp DNA
DEFINITION Sequence 36 from patent US 6342220.
ACCESSION AR183492
VERSION AR183492.1 GI:20227461
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 36 29-JAN-2002;
FEATURES Location/Qualifiers

```

source          1. .15
                /organism="unknown"
BASE COUNT      7 a          4 c          2 g          2 t
ORIGIN
Query Match      100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY              1 AGCCATACATGAC 15
                |||
                1 AGCCATACATGAC 15

Db
1 AGCCATACATGAC 15

RESULT 2
AC024250/c
LOCUS          AC024250          43523 bp          DNA          linear          PRI 09-MAY-2001
DEFINITION    Homo sapiens BAC clone RP11-684N2 from Y, complete sequence.
ACCESSION     AC024250
VERSION       AC024250.6  GI:11120963
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS       Sulston,J.E. and Waterston,R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074
REFERENCE
AUTHORS       Kyung,K., Drone,K. and Le,T.
TITLE         The sequence of Homo sapiens BAC clone RP11-684N2
JOURNAL       Unpublished
REFERENCE
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (28-FEB-2000) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA
4 (bases 1 to 43523)
Waterston,R.
Direct Submission
Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 43523)
Waterston,R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gi:9958320.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0684N02
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

```

```

MAPPING INFORMATION:
The position of this clone was established as part of a
collaboration between the Human Chromosome Y Mapping Project
(Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
and David C. Page at the Whitehead Institute for Biomedical
Research, Cambridge MA) and the Washington University Genome
Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Rateno,M., Catanesse,J.J. and de Jong,P.U. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-943F15; the clone sequenced
to the right is RP11-489O13. Actual start of this clone is at base
position 17386 of RP11-943F15; actual end is at base position 71706
of RP11-489O13.

The sequence fidelity of RP11-684N2 between bases 832 and 43523 can
not be guaranteed due to a tandem repeat. This region contains low
quality, single stranded areas, and base mismatches. Digest
information corresponds to the current assembly.

FEATURES
SOURCE
1. .43523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-684N2"
/clone_lib="RP11-11"
repeat_region 1. .43507
BASE COUNT 9246 a 8789 c 10023 g 15465 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 43523;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY              1 AGCCATACATGAC 15
                |||
                1 AGCCATACATGAC 15

Db              17764 AGCCATACATGAC 17750

RESULT 3
AC079261/c
LOCUS          AC079261          50210 bp          DNA          linear          PRI 09-MAY-2001
DEFINITION    Homo sapiens BAC clone RP11-1325K3 from Y, complete sequence.
ACCESSION     AC079261
VERSION       AC079261.2  GI:11024958
KEYWORDS
ORGANISM      Homo sapiens.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS       Sulston,J.E. and Waterston,R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074
REFERENCE
AUTHORS       Ozerisky,F., Abdoet,A., Maupin,R. and Ames,M.
TITLE         The sequence of Homo sapiens BAC clone RP11-1325K3
JOURNAL       Unpublished
REFERENCE      3 (bases 1 to 50210)

```


translation="MAONEIEDMDLRIRIKSGNIDRVKINRTEKLEMYKHYHLL
PDSFVKKKKRWAKWVWVLDIGIPRECKTNINLPLELHLEFPEGKISLSEYHLL
NDSLSKVMOCVGNKLNVDLITLORAPREPENLEIHRFTKOLKIVOKMFEELYL
VYETNGINYNHEKIECELETNRHHRATVNGOLCALLGYDSNVDGDKDDDEED
EDGDEDEDILNKPPYLLCLITLVELEMKRIFGBOKASFSTOSRTFKDKLKGSH
HLHSLMYLRKKKLEKFPNNNSPONTDIALEFLVFLADAVANHVNGMLNVMKRY
GALGVLYVYOKLLLSIKKDNKISLCSIQWERTKDLKAVENYVSKLEFPTLKIK
EPYVGSFELSLIKLEMSKSCIDFLMKPLGLAGNFMHIEFSLPTLKIK
LRDVAKVEHEHILODLHRTIMAYEAVALDLSILAOYVFMHIEFSLPTLKIK
QINVOYOMMSADIALKPCYVAFEPFPHINVPDEDEIVGENDIMKEMPFLIRG
TNDIDIVPVGSGGKTTVARKVNSDNTVSHQDPAKIVSOTNRRLKLOITISQ
VTGSKDGEEDDILADELRKSMKRLITLDDMOCMADRLISPSGNSRIV
TTRLEKGEYKCHTDPISLPLETESCFLQKRVKQEDFPELQDVSQVAKCK
GLPLVYVAGIKIRKMEESWMLKDFLFDYLDCHSEYSAITVQDLSADLCK
PCLLMGMPEDASITVSALISMLADVPONIESADYLMNLISNVVMSKKEG
KIKCYKVDVYLHFCLEKSRRENFMVAKGNHOSQOFVKOSRVSLSSEENKES
KTRKSHOPLRSLTNGASPDILSLSLNLRLKVLVLSHEDVNEATLKPML
LKYGVARSTYFPDSHPLPLETLYKNNRSVMRGCPWEMEQLRHVSIDAEPRK
CLPESKXENLRLIKNINCPIDRADVLSRGNLOOLHIEFEVAPPMVLTIGLT
OLQILDSFRLSHVSCLOLPNKLMLVLEADDEKLTSTVAGLPSEYIQLTNMFEH
OSEWGLDITFHRKLKLVOLDISRWVDESEFPDLETTITIKMLDEKPLISPDV
PTLKOIKLHCTKSLDSDSAVRIRKIVDEENGNDRLDILIKVSRNKLCLCFVEPYLAS
NLPTLOKWM"
/db_xref="GI:1541873"
/gene="Sw5-d"
/product="L2-NBS-LRR class resistance protein D"
/protein_id="AAG31016.1"
/translation="MAONEIPEMDLRIRIESGNTSSVELEKVEKEMRLRRTIT
RYHRIFPDSLYKTKKAKLIVTLOGLIPDIPRECKTNINLELVSHLEFPOGNTA
SLSCSESDNDVLSBDMWLGKNTNDVILIGQPDNDSCQYRKQIKLHKRREPL
YLATEINGVYDEKLESLERIQMDNVGOLCTAPLVEDEDEDEDELVDLENKP
PYLLGLIVLEKEMKIFLNLKASKFTOSTFKDKLPGFSHLSLVLVYRNKL
KPNPNAQONIGLIEILFELFDLADLSNHYLSNMNLLENVATJAGVILVIOKL
LPRSISKDDSTINITYSMOVLKTKDLKAVETVYKSLFTFPOFPTSGLFEMDL
LRKLNEMRSKYGDFLMPKIPFCKLEKOSTLISNLEKESLSSISIRDTVYHNECK
DLOERTILAYEAGVSIDSCAKYVNMHCCSPILLEETKOLNAVTEMMASDPL
NPHYAAPFKLPTPHRSNPTIDEETIYDEETIYDEKIKIIOCTIRTNDDVPIYG
MGOGKTYIARKVNSDNTVSHQDPAKIVSOTNRRLKLOITISQVTSKDKGID
DILADELRKSLMGKRYLVLDDMDCMADRLISPEDFGNSRIVVTVTVLYVAG
KCHTDPISLPLETESCFLQKRVKQEDFPELQDVSQVAKCKGLPLVYVAG
IIRKMEESWMLDLSLFDYLDCHSEYSAITVQDLSADLCKLPGFSLYVAG
DASIVSALLSLMIAEDVPONIESAEDYLMNLISNVVMSKKEGKIKCYEVDV
LHFCLEKSRRENFMVAKGNHOSQOFVKOSRVSLSSEENKESKTRKSHOPLR
SLTNGASPDILSLSLNLRLKVLVLSHEDVNEATLKPMLVLEADDEKLTSTV
YFDRSHPLPLETLYKNNRSVMRGCPWEMEQLRHVSIDAEPRKISPSGNSRIV
LRLKINIVGPIPDVLSRGNLOOLHIEFEVAPPMVLTIGLTOLQILDSFRLSHV
BRHVSILQLPNKLMLVLEADDEKLTSTVAGLPSEYIQLTNMFEH
DITFHRKLKLVOLDISRWVDESEFPDLETTITIKMLDEKPLISPDVPTLKOIK
LHCKKSLDSDSAVRIRKIVDEENGNDRLDILIKVSRNKLCLCFVEPYLASNLPTLO
K"
/db_xref="GI:1541873"
/gene="Sw5-e"
/product="L2-NBS-LRR class resistance protein E"
/protein_id="AAG31017.1"
/translation="MAONEIPEMDLRIRIESGNTSSVELEKVEKEMRLRRTIT
RYHRIFPDSLYKTKKAKLIVTLOGLIPDIPRECKTNINLELVSHLEFPOGNTA
SLSCSESDNDVLSBDMWLGKNTNDVILIGQPDNDSCQYRKQIKLHKRREPL
YLATEINGVYDEKLESLERIQMDNVGOLCTAPLVEDEDEDEDELVDLENKP
PYLLGLIVLEKEMKIFLNLKASKFTOSTFKDKLPGFSHLSLVLVYRNKL
KPNPNAQONIGLIEILFELFDLADLSNHYLSNMNLLENVATJAGVILVIOKL
LPRSISKDDSTINITYSMOVLKTKDLKAVETVYKSLFTFPOFPTSGLFEMDL
LRKLNEMRSKYGDFLMPKIPFCKLEKOSTLISNLEKESLSSISIRDTVYHNECK
DLOERTILAYEAGVSIDSCAKYVNMHCCSPILLEETKOLNAVTEMMASDPL
NPHYAAPFKLPTPHRSNPTIDEETIYDEETIYDEKIKIIOCTIRTNDDVPIYG
MGOGKTYIARKVNSDNTVSHQDPAKIVSOTNRRLKLOITISQVTSKDKGID
DILADELRKSLMGKRYLVLDDMDCMADRLISPEDFGNSRIVVTVTVLYVAG
KCHTDPISLPLETESCFLQKRVKQEDFPELQDVSQVAKCKGLPLVYVAG
IIRKMEESWMLDLSLFDYLDCHSEYSAITVQDLSADLCKLPGFSLYVAG
DASIVSALLSLMIAEDVPONIESAEDYLMNLISNVVMSKKEGKIKCYEVDV
LHFCLEKSRRENFMVAKGNHOSQOFVKOSRVSLSSEENKESKTRKSHOPLR
SLTNGASPDILSLSLNLRLKVLVLSHEDVNEATLKPMLVLEADDEKLTSTV
YFDRSHPLPLETLYKNNRSVMRGCPWEMEQLRHVSIDAEPRKISPSGNSRIV
LRLKINIVGPIPDVLSRGNLOOLHIEFEVAPPMVLTIGLTOLQILDSFRLSHV
BRHVSILQLPNKLMLVLEADDEKLTSTVAGLPSEYIQLTNMFEH
DITFHRKLKLVOLDISRWVDESEFPDLETTITIKMLDEKPLISPDVPTLKOIK
LHCKKSLDSDSAVRIRKIVDEENGNDRLDILIKVSRNKLCLCFVEPYLASNLPTLO
K"
/db_xref="GI:1541873"
/gene="Sw5-e"
/product="L2-NBS-LRR class resistance protein E"
/protein_id="AAG31017.1"
/translation="MAONEIPEMDLRIRIESGNTSSVELEKVEKEMRLRRTIT
RYHRIFPDSLYKTKKAKLIVTLOGLIPDIPRECKTNINLELVSHLEFPOGNTA
SLSCSESDNDVLSBDMWLGKNTNDVILIGQPDNDSCQYRKQIKLHKRREPL
YLATEINGVYDEKLESLERIQMDNVGOLCTAPLVEDEDEDEDELVDLENKP
PYLLGLIVLEKEMKIFLNLKASKFTOSTFKDKLPGFSHLSLVLVYRNKL
KPNPNAQONIGLIEILFELFDLADLSNHYLSNMNLLENVATJAGVILVIOKL
LPRSISKDDSTINITYSMOVLKTKDLKAVETVYKSLFTFPOFPTSGLFEMDL
LRKLNEMRSKYGDFLMPKIPFCKLEKOSTLISNLEKESLSSISIRDTVYHNECK
DLOERTILAYEAGVSIDSCAKYVNMHCCSPILLEETKOLNAVTEMMASDPL
NPHYAAPFKLPTPHRSNPTIDEETIYDEETIYDEKIKIIOCTIRTNDDVPIYG
MGOGKTYIARKVNSDNTVSHQDPAKIVSOTNRRLKLOITISQVTSKDKGID
DILADELRKSLMGKRYLVLDDMDCMADRLISPEDFGNSRIVVTVTVLYVAG
KCHTDPISLPLETESCFLQKRVKQEDFPELQDVSQVAKCKGLPLVYVAG
IIRKMEESWMLDLSLFDYLDCHSEYSAITVQDLSADLCKLPGFSLYVAG
DASIVSALLSLMIAEDVPONIESAEDYLMNLISNVVMSKKEGKIKCYEVDV
LHFCLEKSRRENFMVAKGNHOSQOFVKOSRVSLSSEENKESKTRKSHOPLR
SLTNGASPDILSLSLNLRLKVLVLSHEDVNEATLKPMLVLEADDEKLTSTV
YFDRSHPLPLETLYKNNRSVMRGCPWEMEQLRHVSIDAEPRKISPSGNSRIV
LRLKINIVGPIPDVLSRGNLOOLHIEFEVAPPMVLTIGLTOLQILDSFRLSHV
BRHVSILQLPNKLMLVLEADDEKLTSTVAGLPSEYIQLTNMFEH
DITFHRKLKLVOLDISRWVDESEFPDLETTITIKMLDEKPLISPDVPTLKOIK
LHCKKSLDSDSAVRIRKIVDEENGNDRLDILIKVSRNKLCLCFVEPYLASNLPTLO
K"

BASE COUNT 19663 a 9668 c 9699 g 20265 t
ORIGIN
Query Match 100.0%; Score 15; DB 8; Length 59465;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCATACATGAC 15
|||||
Db 50154 AGCCATACATGAC 50168
RESULT 5
AC106140/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-230F16, *** SEQUENCING IN PROGRESS
*** 54 unordered pieces.
AC106140
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Norway rat.
Rattus norvegicus
Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS
1 (bases 1 to 91240)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,F.,
Barbarta,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,
Bouay,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Datborne,S.R., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinn,H.H.,
Douthwaite,K.J., Drepper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Haje,S., Hamilton,K.,
Harris,C., Harries,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hughes,M., Hollaway,C., Hollins,B.,
Homs,I.E., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Kapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzkes,M.,
Moser,M., Neal,D., Newton,J., Newton,M., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,M., Morgan,M., Morris,S.,
Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okunonu,G.,

Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, C.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Soderren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Sytek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 91240)
Worley, K.C.

Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 91240)
Worley, K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18138661.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GJGR
Center clone name: CH230-230F16

Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 36346 bases at least Q40
Consensus quality: 39251 bases at least Q30
Consensus quality: 42076 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(use http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1264: contig of 1264 bp in length
* 1265 1364: gap of unknown length
* 1365 2650: contig of 1286 bp in length
* 2651 2750: gap of unknown length
* 2751 3941: contig of 1191 bp in length
* 3942 4041: gap of unknown length
* 4042 5286: contig of 1245 bp in length
* 5287 5386: gap of unknown length
* 5387 6670: contig of 1284 bp in length
* 6671 7909: contig of 1139 bp in length
* 7910 8009: gap of unknown length
* 8010 9162: contig of 1153 bp in length
* 9163 9262: gap of unknown length
* 9263 10570: contig of 1308 bp in length
* 10571 10670: gap of unknown length
* 10671 11724: contig of 1054 bp in length
* 11725 11824: gap of unknown length
* 11825 12998: contig of 1174 bp in length
* 12999 13098: gap of unknown length

13099 14265: contig of 1167 bp in length
* 14266 14365: gap of unknown length
* 14366 15589: contig of 1224 bp in length
* 15590 15689: gap of unknown length
* 15690 16720: contig of 1031 bp in length
* 16721 16820: gap of unknown length
* 16821 17903: contig of 1083 bp in length
* 17904 18003: gap of unknown length
* 18004 19222: contig of 1219 bp in length
* 19223 19323: gap of unknown length
* 19323 20454: contig of 1132 bp in length
* 20455 20554: gap of unknown length
* 20555 21851: contig of 1297 bp in length
* 21852 21951: gap of unknown length
* 21952 23916: contig of 1964 bp in length
* 23916 24015: gap of unknown length
* 24016 25805: gap of unknown length
* 25806 25905: gap of unknown length
* 25906 27164: contig of 1258 bp in length
* 27164 27263: gap of unknown length
* 27264 28571: contig of 1308 bp in length
* 28572 28672: gap of unknown length
* 28672 30050: contig of 1379 bp in length
* 30051 30150: gap of unknown length
* 30151 31772: contig of 1621 bp in length
* 31772 31872: gap of unknown length
* 31872 33132: contig of 1261 bp in length
* 33133 33232: gap of unknown length
* 33233 34592: contig of 1360 bp in length
* 34593 34692: gap of unknown length
* 34693 35740: contig of 1048 bp in length
* 35741 37229: gap of unknown length
* 37229 37329: contig of 1389 bp in length
* 37329 38840: gap of unknown length
* 38840 38941: contig of 1511 bp in length
* 38941 40795: gap of unknown length
* 40795 40895: contig of 1855 bp in length
* 40896 43165: gap of unknown length
* 43165 43265: contig of 2270 bp in length
* 43266 44613: gap of unknown length
* 44614 44713: contig of 1348 bp in length
* 44713 45767: gap of unknown length
* 45767 45768: contig of 1054 bp in length
* 45768 47485: gap of unknown length
* 47485 47585: contig of 1618 bp in length
* 47586 49220: gap of unknown length
* 49220 49320: contig of 1635 bp in length
* 49321 49321: gap of unknown length
* 49321 50434: contig of 1114 bp in length
* 50435 50334: gap of unknown length
* 50335 52376: contig of 1842 bp in length
* 52376 52475: gap of unknown length
* 52477 54365: contig of 1889 bp in length
* 54366 54465: gap of unknown length
* 54466 55482: contig of 1017 bp in length
* 55483 55582: gap of unknown length
* 55583 57326: contig of 1744 bp in length
* 57327 57426: gap of unknown length
* 57427 58482: contig of 1056 bp in length
* 58483 60390: gap of unknown length
* 60390 60490: contig of 1808 bp in length
* 60491 62642: gap of unknown length
* 62643 62742: contig of 2152 bp in length
* 62743 64159: gap of unknown length
* 64159 64259: contig of 1417 bp in length
* 64260 64259: gap of unknown length
* 64260 65435: contig of 1176 bp in length
* 65436 65335: gap of unknown length
* 65336 68178: contig of 2643 bp in length
* 68179 71375: gap of unknown length
* 71376 71475: contig of 3097 bp in length
* 71476 73359: gap of unknown length
* 73359 73359: contig of 1884 bp in length

	*	73360	73459:	gap of unknown length
	*	73460	75813:	contig of 2354 bp in length
	*	75814	75913:	gap of unknown length
	*	75914	77801:	contig of 1888 bp in length
	*	77802	77901:	gap of unknown length
	*	77902	79798:	contig of 1897 bp in length
	*	79799	79898:	gap of unknown length
	*	79899	82762:	contig of 2864 bp in length
	*	82763	82862:	gap of unknown length
	*	82863	86500:	contig of 3538 bp in length
	*	86501	86500:	gap of unknown length
	*	86501	88336:	contig of 1836 bp in length
	*	88337	88436:	gap of unknown length
	*	88437	91240:	contig of 2804 bp in length.
Query Match		100.0%:	Score 15:	DB 2: Length 91240:
Best Local Similarity	100.0%:	Pred. No. 2e+02:		
Matches 15:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	1	AGCCATACATGAAC	15	
Db	88261	AGCCATACATGAAC	88247	
RESULT 6				
AC023754		100685 bp	DNA	linear
LOCUS	Arabidopsis thaliana chromosome I BAC F1B16			genomic sequence,
DEFINITION	complete sequence.			
AC023754				
VERSION	AC023754.3	GI:7212003		
KEYWORDS	HTG.			
SOURCE	Arabidopsis thaliana.			
ORGANISM	Arabidopsis thaliana.			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.			
REFERENCE	1	(bases 1 to 100685)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,			
	Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bel,O., Buehler,E.,			
	Buehler,E., Chao,Q., Chin,C., Chou,J., Choi,E., Gonzalez,A.,			
	Howing,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,J., Lee,J.M.,			
	Lenz,C., Liu,A., Liu,S., Mukharisky,N., Pham,P., Sakano,H.,			
	Shin,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A.			
	and Davis,R.W.			
	Unpublished			
JOURNAL	2	(bases 1 to 100685)		
REFERENCE	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,			
AUTHORS	Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bel,O., Buehler,E.,			
	Chin,C., Chou,J., Choi,E., Dunn,P., Gonzalez,A., Howing,B., Kim,C.,			
	Koo,J., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,			
	Mukharisky,N., Pham,P., Sakano,H., Schwartz,J., Shin,P.,			
	Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,			
	Theologis,A. and Davis,R.W.			
	Direct Submission			
TITLE	Submitted (17-FEB-2000) DNA Sequencing and Technology Center,			
JOURNAL	Stanford University, 855 California Avenue, Palo Alto, CA 94304,			
	USA			
REFERENCE	3	(bases 1 to 100685)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,			
	Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bel,O., Buehler,E.,			
	Chin,C., Chou,J., Choi,E., Dunn,P., Gonzalez,A., Howing,B., Kim,C.,			
	Koo,J., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,			
	Mukharisky,N., Pham,P., Sakano,H., Schwartz,J., Shin,P.,			
	Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,			
	Theologis,A. and Davis,R.W.			
	Direct Submission			
TITLE	Submitted (09-MAR-2000) DNA Sequencing and Technology Center,			
JOURNAL	Stanford University, 855 California Avenue, Palo Alto, CA 94304,			
	USA			
REFERENCE	4	(bases 1 to 100685)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,			
	Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.			

and Davis,R.W.
Direct Submission
Submitted (13-SEP-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

On Mar 9, 2000 this sequence version replaced gi:7143418.
Bases 132,040 of IGF clone F1B16 overlap with bases 97,508-129,547
of IGF clone F10A5, gb|AC006434.
e-mail for correspondence: arab@sequence.stanford.edu
Genes with similarity to proteins in the databases are named
'putative', '-like' or 'similar to'. Genes that have EST
similarity but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M.
Hebbsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).

Location/Qualifiers
1..100685
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="I"
/clone="F1B16"
1..32040
/note="overlap with bases 97,508-129,547 of IGF clone
F10A5, gb|AC006434. See Genbank record for BAC clone
F10A5 for annotation in this region."
complement(32986..33822)
/gene="F1B16.1"
complement(32986..33822)
/gene="F1B16.1"
complement(32986..33822)
/note="Unknown protein; unknown protein"
/codon_start=1
/protein_id="AA013069.1"
/db_xref="GI:10120444"
/translation="MALPIVSPSSSTSLSPFSKSPPHSLRSLSPYDNRCKLN
SSFRSSSSSEKHNNAKSPKSDIVLEPLPLPLFGAILPLDIEFFRRIMMT
LIQSLDRFQVNVSVSGSAAEVGVGVEVKKERLVDDEFLVCGOEFRTNVVNT
KPYLVGEVFWLLEDRSGEENLDSLNEVEVLMKEVIRLSNRINGKAERKVDLRNPF
RPVSPFVSTFEGCAPREQDALLLEDDTAARLKRERETLRNTLNTLTAASAKVDPPS
SS"
43756..46560
/gene="F1B16.2"
join(43756..44347,44507..44562,44700..44966,45493..45755,
46200..46560)
/gene="F1B16.2"
/note="Similar to cytokinin oxidase"
/codon_start=1
/protein_id="AA013068.1"
/db_xref="GI:10120443"
/translation="MTSSFLLTFAICKLIIVAGLNVGSPSELLRIGALIDVDSHTVHP
SDLASVSDFGMLKSPDEPLAVLHPSSAEVRAVRYAGSATAPVARGHGSHNG
OAAAGRNCGVEMNGGCTPKPIPRPEMDYDVWGGLMDVYLKTEHGLAPKSWT
DYLTVTNGTISNAGISQAFRHGQKJQISNVLVDVYTOFGITTRARSLEAPQVRI
MIRLVYSSFKVFTEDQEVLTISMHGQKJQISNVLVDVYTOFGITTRARSLEAPQVRI
VSSKSGSLVLCLEITKMYHDSDEIVDQVEIILKKKLNFIPTSVFTDLOYDFLRI
VHKAELKRLKRLMEVPPHWNLEVPKRSISDFDKGVKGLIGNTSGPILLYPNKD
KDKERSAVTDEDEVEFYLVALLRSALTGEEROKLEYLKDNRRLTFECDEAKTVKO
YLPHTATQEWVAHGDMDKDFRSIKAFEDFRHILATQGRIFQNPNSLSLPPSSSSSS
AASW"
complement(46848..48185)
/gene="F1B16.3"
complement(join(46848..46943,47040..47128,47474..47537,
47773..47875,47963..48066,48156..48185))
/gene="F1B16.3"
/codon_start=1
/product="Putative ubiquitin-conjugating enzyme"
/protein_id="AA013066.1"
/db_xref="GI:10120441"

```
gene
52291..53501
/gene="F1B16.4"
join(52291..52584,52917..53501)
/gene="F1B16.4"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AA013063.1"
/db_xref="GI:10120438"
/translation="MEDPVRHEGSSLRGLTLDNRAYAKAVOCLOVEVIDIGREVEELC
NNILINOLPGRRRPPGAFSSRIKSELSSGFMSLPENHEIHKITKLLSLLOOYER
FEOYCNOLBOVISFEETIAGBESSKYVTGLAIQAMTRHRSIEBAITQOLNVRERF
ISHODPKITSSGLSOLSLDENTSSSLGRLGLOVGPORHAKTIRGLPETSAVILR
AMLFOHFLHPSVLEFKCNNTSDESKLCKMWNNTSLIAVMKTVIRMKORSCWRLKOD
FRRTKYTPNLILFF"
complement(54491..56307)
/gene="F1B16.5"
complement(join(54491..55074,55197..55351,55498..55769,
55872..56014,56253..56307))
/gene="F1B16.5"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AA013070.1"
/db_xref="GI:10120445"
/translation="KGTDFGVYSICINSYDSRGLLMEFLARGVADYVNTNO
KPLEDEVVLSLEHKMLDRGOVLSAKGKAVDTSLKALILAVNTRAKWLDAYIKM
NVKVLPLIMVHEMRGHFNADLVKHLPEVAGAIIDSHADIAVNTQARLGIKM
PRTYVVLGNSLEMAEDSVAKRVLREHVESLGEVRNEDLLFGLINSVSGKODL
FLAPHESLERIEKKEKLOVPMHAAVVGSDMSKOTETLENFVREKELNEVHFVN
KLTIVAPYIAIDVLQNSOARGECGRITTEEMARKFLVLGTAAGVGTIYVNGTGG
LHSAKREKVEYPLAKNIVLATVOVELRLMKNGRYRVAEMPLEHMSRISVLYKEV
LQHKARALR"
59918..61713
/gene="F1B16.6"
join(59918..60655,60734..61110,61188..61248,61315..61713)
/gene="F1B16.6"
/notes="Similar to homeodomain proteins"
/codon_start=1
/protein_id="AA013065.1"
/db_xref="GI:10120440"
/translation="MAVYYPNSVGMOSLYOEIYIENEOOQOQOQSSASPSSEIVS
GVRNNMEYIPTSDVAVNGANTVSSNDLSFHGGSLSLGNOISAVSFPQYHE
ONISLQSNININPTSDMDEKSLSVHOHSDOILPSSYNNNGNNGGVFNRYE
TSGFVSVALRSRYLKEPOOLDEYVSVRKDKIKMKMKNDKODPHNGSSNITPDD
KTSOELSPERLOSKSKSLTLDVDEVDKRTNYTHQNEALASSFEWYTGIAKAP
KYTSVALNRISRHFRCLDAIKQIYIRKGLERETSDQGERIPRLRYLDORLROO
ALHQOLGAVRPAWRPQGLPENSVSILRAMLPFHLHPYKSEKIMLSKOTGLSNQ
VAMFPIINARVRLKPMIEEMYEKEEGSEALLSNODTKMKQETSQLEHDESSSQO
ONQNNNNNINPTSDAONLVFADPKPRDATTGDYDSLMWYHGFIDYRNYGLCNO
QDRYSNPQHLHDFV"
complement(62425..64698)
/gene="F1B16.7"
complement(join(62425..63429,63528..63559,63863..64074,
64640..64698))
/gene="F1B16.7"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AA013067.1"
/db_xref="GI:10120442"
/translation="MPFRIVSCDFPPRNLSLFMRVADIEGSPYSPYVAVRQISM
DKMSLGERGTLSEGGTPDGHGTPASOKATPEMSTNMVPPSSGSLASSDLIEV
KNLYDSPIGVSAOSKPLSTPLPNVPSLSSAHRILPKTPSPRARSRQDLDF
ROYSDSOIILKSPKINNYSTSEGRSSFTVSTCSDNATGTSVSSGCGMFMFSLY
ATISORDWSTFDSHILSGRRKLSGSSRFSPSYVDVOYCGACSKILTERSSIAFE
LPIAAVLACGHVYHAECLFTMTDIEKYDACPCTTGEKRVAKITRKALKAAEAKA
KOYKRCNRRVSDYGESECEDFVFOKMGKREGALKLEACSSSKNSKSLKMFHPS
ISSKWNKPSKSDALKKGFWRRNRSSSSSIEVKIHTLIQS"
complement(69063..69566)
/gene="F1B16.8"
complement(69063..69566)
```

```

Query Match
Best Local Similarity 100.0% Score 15; DB 8; Length 100685;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCTACATGAC 15
DB 16468 AGCCTACATGAC 16482

RESULT 7
AC121474
LOCUS
DEFINITION
Rattus norvegicus clone CH230-88W5, *** SEQUENCING IN PROGRESS ***
62 unordered pieces.
AC121474
AC121474.2 GI:21909158
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 101282)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-oman,F.R., Allen,C.,
Alsdrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Blimang,K., Blankenburg,K., Bonnin,D.,
Boucek,J., Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Dreper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgall,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,T., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Gatz,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,J.F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudan,S.,
Karissom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kraivolic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Wei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwu,S., Ogih,M., Okuwonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Prims,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scheer,S., Scott,G., Shen,H., Shooshari,N., Sisson,J.,
Sutougan,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Tamerisa,K., Tang,H.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Wotley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
TITLE
JOURNAL
Direct Submission
Unpublished
```

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 101282)
Worley, K. C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 101282)
Worley, K. C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:20976426.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GYV
Center clone name: CH230-88M5

Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 53833 bases at least Q40
Consensus quality: 58211 bases at least Q30
Consensus quality: 60723 bases at least Q20.

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1095: contig of 1095 bp in length
1195: gap of unknown length
1196
2364: contig of 1169 bp in length
2365
2464: gap of unknown length
2465
3475: contig of 1011 bp in length
3476
3575: gap of unknown length
3576
4593: contig of 1018 bp in length
4594
4693: gap of unknown length
4694
5854: contig of 1161 bp in length
5855
5954: gap of unknown length
5955
6984: contig of 1030 bp in length
6985
7084: gap of unknown length
7085
8190: contig of 1106 bp in length
8191
8290: gap of unknown length
8291
9708: contig of 1418 bp in length
9709
9808: gap of unknown length
9809
11270: contig of 1462 bp in length
11271
11370: gap of unknown length
11371
12681: contig of 1311 bp in length
12682
12781: gap of unknown length
12783
13856: contig of 1075 bp in length
13857
13957: gap of unknown length
13958
15348: contig of 1392 bp in length
15349
15448: gap of unknown length
15449
16775: contig of 1327 bp in length
16776
16875: gap of unknown length
16876
18345: contig of 1470 bp in length
18346
18445: gap of unknown length
18446
19882: contig of 1437 bp in length
19883
19982: gap of unknown length
19983
21546: contig of 1564 bp in length
21547
21646: gap of unknown length
21647
23246: contig of 1600 bp in length
23247
23346: gap of unknown length

23347
24676: contig of 1330 bp in length
24677
24776: gap of unknown length
24777
26040: contig of 1264 bp in length
26041
26140: gap of unknown length
26141
27293: contig of 1153 bp in length
27294
27394: gap of unknown length
27395
28872: contig of 1479 bp in length
28873
28973: gap of unknown length
28974
30187: contig of 1215 bp in length
30188
30287: gap of unknown length
30288
31484: contig of 1197 bp in length
31485
31584: gap of unknown length
31585
32506: contig of 1022 bp in length
32507
32706: gap of unknown length
32707
33988: contig of 1282 bp in length
33989
34088: gap of unknown length
34089
35495: contig of 1407 bp in length
35496
35595: gap of unknown length
35596
37634: contig of 2039 bp in length
37635
37734: gap of unknown length
37735
39181: contig of 1447 bp in length
39182
39281: gap of unknown length
39282
40774: contig of 1493 bp in length
40775
40874: gap of unknown length
40875
42029: contig of 1155 bp in length
42030
42129: gap of unknown length
42130
43727: contig of 1598 bp in length
43728
43827: gap of unknown length
43828
45082: contig of 1255 bp in length
45083
45182: gap of unknown length
45183
46312: contig of 1130 bp in length
46313
46413
47654: gap of unknown length
47655
47754: gap of unknown length
47755
48861: contig of 1107 bp in length
48862
48961: gap of unknown length
48962
50268: contig of 1307 bp in length
50269
50368
50369
51686: gap of unknown length
51687
51786: contig of 1318 bp in length
51787
52914: gap of unknown length
52915
53014: contig of 1128 bp in length
53015
54429: gap of unknown length
54430
54529: contig of 1415 bp in length
54530
55937: gap of unknown length
55938
56037: contig of 1408 bp in length
56038
57643: gap of unknown length
57644
57749: contig of 1612 bp in length
57750
59809: gap of unknown length
59810
59909: contig of 2060 bp in length
59910
61896: gap of unknown length
61897
61966: contig of 1987 bp in length
61967
63329: gap of unknown length
63329
63429: contig of 1333 bp in length
63430
64748: gap of unknown length
64749
64848: contig of 1319 bp in length
64849
66048: gap of unknown length
66049
66148: contig of 1200 bp in length
66149
67773: gap of unknown length
67774
67873: contig of 1625 bp in length
67874
69469: gap of unknown length
69470
69569: contig of 1596 bp in length
69570
71241: gap of unknown length
71242
71341: contig of 1672 bp in length
71342
71343: gap of unknown length
71344
73081: gap of unknown length
73082
73181: contig of 1740 bp in length
73182
75302: gap of unknown length
75303
75402: contig of 2121 bp in length
75403
77624: gap of unknown length
77624
77724: contig of 2222 bp in length
77725
79324: gap of unknown length
79324
79424: contig of 1600 bp in length
79425
80888: gap of unknown length
80888
81464: contig of 1464 bp in length

Query Match	100.0%;	Score 15;	DB 2;	Length 101282;
Best Local Similarity-	100.0%;	Pred. No. 2e+02;		
Matches 15; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0

QY 1 AGCCATAACATGAAC 15
|||||
Db 86816 AGCCATAACATGAAC 86830

RESULT 8	
LOCUS	AC009240/c
DEFINITION	AC009240 Homo sapiens BAC clone Rpl1-489013 from Y, complete sequence.
ACCESSION	AC009240
VERSION	AC009240.6
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 101290)
AUTHORS	Sulston, J.E. and Waterston, R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
PUBMED	9847074

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
2	(bases 1 to 101290)	Scott, K., Holmes, A., Maupin, R., and Hodges, J.		
		The sequence of Homo sapiens BAC clone RP11-4690J3		
		unpublished		
3	(bases 1 to 101290)			

JOURNAL
Submitted (06-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
4 (bases 1 to 101230)
Waterston, R.H.
Direct Submission
Submitted (15-SEP-2000) Genome Sequencing Center, Washington

REFERENCE 5 (bases 1 to 101290)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 15, 2000 this sequence version replaced g1:9838226.
----- genome center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@watsn.wustl.edu
----- Summary Statistics -----
Center project name: H_NH048013

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a

collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RpCl-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.-Y., Zhao, B., Frengen, E., Tatematsu, M., Catalanese, J.J., and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACet.6

The clone sequenced to the left is RP11-684N2; the clone sequenced to the right is RP11-329C15, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-489013; actual end is at base position 73076 of RP11-329C15.

The sequence H₂NH0489013 from base position 1 to 32763 contains a tandem repeat which continues into Rp11-684N2. The sequence fidelity of the tandem repeat cannot be guaranteed. Digest data was used to make the assembly however, it still suggests there is missing data.

```
FEATURES      Location/Qualifiers
source        1. .101290
```

```
/organism="Homo sapiens"  
/db xref="taxon:9606"
```

```

/chromosome="Y"
/map="Y"

```

```
/map - 1
/clone="RP11-489013"
```

```
repeat_region 18. .32348 /clone_110=RPC
```

```
repeat_region      /rpt_family="E" 32385. .32820
```

```

repeat_region 32810 33021 /rpt_family="L1

```

```
repeat_region 32010 1 32022  
/rpt_family="L1  
32046 34646
```

```
repeat_region 33040 : 33040
/rpt_family="L1
```

```
repeat_region 34914..35140
/rpt_family="A1
```

```
repeat_region 35144. 36404
/rnt familyv="L1
```

repeat_region	36408.37075
---------------	-------------

```
repeat_region      37076.  .37374
```

```
repeat_region 37375..38180 /rpt_family="A1
```

```
repeat region 38191 38328
/rpt_family="L1
```

```
report_location  
/rpt_family="L"  
3956A 39825
```

```
repeatL_region 36004 : 36033  
/rpt_family="L"
```

```
repeat_region 39841..39894
/rpt_family="A.
```

```
repeat_region 39920.40231
/rpt_family="A"
```

```
repeat_region 40388.40988
```

```
repeat_region 40990. 41925
```

```
repeat_region      /rpt_family="L.  
41981.    .42074
```

```

/rpt_family="L
42075. .42369
Repeat region

```

```

report region
42370 42630
/rpt_family="A

```

```
repeat_region 42370: 42400  
/rpt_family="L 42400 42400
```

```
repeat_region 43200..43400
/rpt_family="L
```

```

repeat_region      43485..43853
                    /rpt_family="L1"
repeat_region      43879..44034
                    /rpt_family="L1"
repeat_region      44421..44673
                    /rpt_family="L1"
repeat_region      44876..45175
                    /rpt_family="L1"
repeat_region      45939..46149
                    /rpt_family="L1"
repeat_region      46241..46473
                    /rpt_family="L1"
repeat_region      46474..46769
                    /rpt_family="L1"
repeat_region      46770..46945
                    /rpt_family="L1"
repeat_region      47026..47143
                    /rpt_family="L1"
repeat_region      47158..47345
                    /rpt_family="L1"
repeat_region      47891..48318
                    /rpt_family="L1"
repeat_region      48332..48394
                    /rpt_family="L1"
repeat_region      48773..49076
                    /rpt_family="L1"
repeat_region      55508..55666
                    /rpt_family="L1"
repeat_region      55675..56009
                    /rpt_family="L1"
repeat_region      56426..56710
                    /rpt_family="L1"
repeat_region      56753..57206
                    /rpt_family="L1"
repeat_region      59131..59503
                    /rpt_family="L1"
repeat_region      59810..59935
                    /rpt_family="L1"
repeat_region      59949..59990
                    /rpt_family="L1"
repeat_region      60011..60189
                    /rpt_family="L1"
repeat_region      60294..60498
                    /rpt_family="L1"
repeat_region      60502..61724
                    /rpt_family="L1"
repeat_region      61725..63600
                    /rpt_family="L1"
repeat_region      63601..67996
                    /rpt_family="L1"
repeat_region      68032..68464
                    /rpt_family="L1"
repeat_region      68468..69183
                    /rpt_family="L1"
repeat_region      69239..69529
                    /rpt_family="L1"
repeat_region      69737..70017
                    /rpt_family="L1"
repeat_region      70050..70099
                    /rpt_family="L1"
repeat_region      71896..73699
                    /rpt_family="L1"
repeat_region      73726..73887
                    /rpt_family="L1"
repeat_region      73998..74248
                    /rpt_family="L1"

```

Query Match Best Local Similarity 100.0%; Score 15; DB 9; Length 101290;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCATTAACATGAAC 15
|||||

Db 12684 AGCCATTAACATGAAC 12670

RESULT 9
AC105725
LOCUS
DEFINITION
AC105725.2 108610 bp DNA linear HTG 13-JUL-2002
Rattus norvegicus clone CH230-64E5, *** SEQUENCING IN PROGRESS ***
56 unordered pieces.
AC105725
VERSION
KEYWORDS
HTG: HTGS PHASE1.
SOURCE
ORGANISM
Rattus norvegicus

REFERENCE
AUTHORS
1 (bases 1 to 108610)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayalew,M., Banks,T.,
Barbata,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Datorne,S.R., David,R.,
Dayila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisl,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunatane,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisged,H.,
Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Madeshwarl,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogdu,M., Okwonnun,G.,
Oragunye,N., Oyaleto,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Pitmus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 108610)

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 108610)

COMMENT
On Jul 12, 2002 this sequence version replaced gi:18092948.
----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GNCM
 Center clone name: CH230-64F5
 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 67703 bases at least Q40
 Consensus quality: 72041 bases at least Q30
 Consensus quality: 75764 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 56 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1010:	contig of 1010 bp in length
1011	1110:	gap of unknown length
1111	2528:	contig of 1418 bp in length
2529	2628:	gap of unknown length
2629	4257:	contig of 1629 bp in length
4258	4357:	gap of unknown length
4358	5571:	contig of 1214 bp in length
5572	5671:	gap of unknown length
5672	6914:	contig of 1243 bp in length
6915	7014:	gap of unknown length
7015	8117:	contig of 1103 bp in length
8118	10026:	contig of 1809 bp in length
10027	10126:	gap of unknown length
10127	11596:	contig of 1470 bp in length
11597	11696:	gap of unknown length
11697	13146:	contig of 1450 bp in length
13147	13246:	gap of unknown length
13247	14748:	contig of 1502 bp in length
14749	14848:	gap of unknown length
14849	16421:	contig of 1573 bp in length
16422	16521:	gap of unknown length
16522	18572:	contig of 2051 bp in length
18573	18672:	gap of unknown length
20389	20389:	contig of 1717 bp in length
20480	21514:	gap of unknown length
21515	21614:	gap of unknown length
21615	22642:	contig of 1028 bp in length
22643	22742:	gap of unknown length
22743	23756:	contig of 1014 bp in length
23757	23856:	gap of unknown length
23857	24874:	contig of 1018 bp in length
24875	24974:	gap of unknown length
24975	26410:	contig of 1436 bp in length
26411	26510:	gap of unknown length
28869	28869:	contig of 2359 bp in length
28870	28969:	gap of unknown length
28970	30219:	contig of 1250 bp in length
30220	30319:	gap of unknown length
30320	32382:	contig of 2063 bp in length
32383	32482:	gap of unknown length
32483	34188:	contig of 1706 bp in length
34189	34288:	gap of unknown length
34289	36188:	contig of 1900 bp in length
36189	36288:	gap of unknown length
36289	38198:	contig of 1810 bp in length
38199	38298:	gap of unknown length

38299	39999:	contig of 1701 bp in length
40000	40099:	gap of unknown length
40100	41725:	contig of 1626 bp in length
41726	41825:	gap of unknown length
41826	43012:	contig of 1187 bp in length
43013	43112:	gap of unknown length
43113	44380:	contig of 1268 bp in length
44381	44480:	gap of unknown length
44481	46394:	contig of 1914 bp in length
46395	46494:	gap of unknown length
46495	47880:	contig of 1386 bp in length
47881	47980:	gap of unknown length
47981	50212:	contig of 2232 bp in length
50213	50312:	gap of unknown length
50313	52820:	contig of 2508 bp in length
52821	52920:	gap of unknown length
52921	54278:	contig of 1358 bp in length
54279	54378:	gap of unknown length
54379	55889:	contig of 1511 bp in length
55890	55989:	gap of unknown length
55990	57668:	contig of 1679 bp in length
57669	57768:	gap of unknown length
57769	60060:	contig of 2292 bp in length
60061	60160:	gap of unknown length
60161	62005:	contig of 1845 bp in length
62006	62105:	gap of unknown length
62106	63410:	contig of 1305 bp in length
63411	63510:	gap of unknown length
63511	64895:	contig of 1385 bp in length
64896	64995:	gap of unknown length
64996	67080:	contig of 2085 bp in length
67081	67180:	gap of unknown length
67181	68364:	contig of 1184 bp in length
68365	68464:	gap of unknown length
68465	71000:	contig of 2536 bp in length
71001	71100:	gap of unknown length
71101	72970:	contig of 1870 bp in length
72971	73070:	gap of unknown length
73071	74831:	contig of 1761 bp in length
74832	74931:	gap of unknown length
74932	76508:	contig of 1577 bp in length
76509	76608:	gap of unknown length
76609	78985:	contig of 2377 bp in length
78986	79085:	gap of unknown length
79086	81746:	contig of 2661 bp in length
81747	81846:	gap of unknown length
81847	84951:	contig of 3105 bp in length
84952	85051:	gap of unknown length
85052	87753:	contig of 2702 bp in length
87754	87853:	gap of unknown length
87854	90939:	contig of 3086 bp in length
90940	91039:	gap of unknown length
91040	94370:	contig of 3331 bp in length
94371	94470:	gap of unknown length
94471	96108:	contig of 1658 bp in length
96109	96208:	gap of unknown length
96209	98946:	contig of 2738 bp in length
98947	99046:	gap of unknown length
99047	101632:	contig of 2586 bp in length

Query Match 100.0%; Score 15; DB 2; Length 108610;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATACATGAAC 15
 |||||||||||||||
 Db 9039 AGCCATACATGAAC 9053

RESULT 10
 H5135122/C
 LOCUS H5135122 121212 bp DNA linear PRI 17-MAR-2000
 DEFINITION Human DNA sequence from clone RPL-135L22 on chromosome 6p22.2-22.3.

Contains STSs and GSSs, complete sequence.

ACCESSION AL031767
VERSION AL031767.13 GI:6969156
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 121212)
AUTHORS Smalley, C.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Feb 11, 2000 this sequence version replaced gi:6911919.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP1-135L22 is from the library RPCT-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-135L22.

FEATURES
SOURCE
Location/Qualifiers
1..121212
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="P22.2-22.3"
/clone="RP1-135L22"
/clone_lib="RPCT-1"
578..853
/note="Alusx repeat: matches 21..286 of consensus"
899..1064
/note="L1MB2 repeat: matches 6001..6166 of consensus"
2374..2639
/note="Alusx repeat: matches 1..280 of consensus"
2976..4675
/note="L1PA7 repeat: matches 4444..6143 of consensus"
4748..4900
/note="MIR repeat: matches 46..221 of consensus"
4975..5128
/note="MER63 repeat: matches 671..1681 of consensus"
5388..5871
/note="MER39 repeat: matches 54..543 of consensus"
5530..6001
/note="match: GSS: Em:B48646"
5947..6269
/note="Alusx repeat: matches 1..312 of consensus"
6994..7284
/note="Alusx repeat: matches 1..291 of consensus"
8006..8317
/note="Alusx repeat: matches 1..313 of consensus"
8425..8721
/note="Alusx repeat: matches 1..306 of consensus"

9185..9646
/note="Tigger3b repeat: matches 757..1229 of consensus"
9648..9689
/note="21 copies 2 mer tc 78% conserved"
9738..10018
/note="Alusx repeat: matches 1..279 of consensus"
10019..10443
/note="Tigger3b repeat: matches 351..767 of consensus"
10444..10739
/note="Alusx repeat: matches 1..296 of consensus"
10740..11066
/note="Tigger3b repeat: matches 1..351 of consensus"
11105..11249
/note="L2 repeat: matches 2605..2750 of consensus"
11257..11713
/note="match: GSS: Em:AQ187585"
11279..11639
/note="L1MC/D repeat: matches 5366..5739 of consensus"
11730..12003
/note="L1R23 repeat: matches 1..278 of consensus"
12004..12299
/note="Alusx repeat: matches 1..294 of consensus"
12300..12326
/note="L1R23 repeat: matches 278..303 of consensus"
12332..12630
/note="Alusx repeat: matches 1..302 of consensus"
12647..12953
/note="Alusx repeat: matches 1..305 of consensus"
12957..13088
/note="L1R23 repeat: matches 299..429 of consensus"
13175..13601
/note="MER39 repeat: matches 13..477 of consensus"
14084..14389
/note="Alusx repeat: matches 1..305 of consensus"
14477..14767
/note="Alusx repeat: matches 1..290 of consensus"
14864..14999
/note="MIR repeat: matches 109..244 of consensus"
16325..16543
/note="Alur repeat: matches 94..311 of consensus"
17085..17198
/note="MIR repeat: matches 70..186 of consensus"
18083..18283
/note="Alur repeat: matches 96..296 of consensus"
19807..19850
/note="22 copies 2 mer tt 77% conserved"
19980..20284
/note="Alur repeat: matches 1..304 of consensus"
20790..21099
/note="M1R1A2 repeat: matches 3..312 of consensus"
21109..21302
/note="97 copies 2 mer tt 57% conserved"
21445..21719
/note="Alur repeat: matches 13..295 of consensus"
21720..21791
/note="M1R1A2 repeat: matches 323..394 of consensus"
22083..22510
/note="M1R1C repeat: matches 6..466 of consensus"
22671..23511
/note="L1PB2 repeat: matches 5254..6154 of consensus"
24714..24745
/note="16 copies 2 mer aa 93% conserved"
25239..25393
/note="MIR repeat: matches 28..188 of consensus"
25468..25690
/note="MER58A repeat: matches 1..221 of consensus"
27578..27688
/note="M1R1B repeat: matches 268..390 of consensus"
27694..28261
/note="L1PA14 repeat: matches 5500..6149 of consensus"
28538..30098
/note="L1PA7 repeat: matches 4586..6141 of consensus"
30256..30567

misc_feature /note="assembly_name:Contig14"
113906..115892
/note="assembly_name:Contig6"
misc_feature 115993..117914
/note="assembly_name:Contig7
clone_end:SP6
vector_side:right"
misc_feature 118015..122126
/note="assembly_name:Contig8"
misc_feature 122327..127400
/note="assembly_name:Contig9"
BASE COUNT 36146 a 27561 c 26883 g 36009 t 801 others
ORIGIN
Query Match 100.0% Score 15; DB 2; Length 127400;
Best Local Similarity 100.0%; Pred. NO. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCATAACATGAC 15
|||||
Db 63666 AGCCATAACATGAC 63652
RESULT 12 129547 bp DNA linear PLN 25-JUL-2000
AC006434
LOCUS
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F10A5, complete
ACCESSION AC006434
VERSION AC006434.5 GI:4662609
KEYWORDS HMG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 129547)
Ecker,J.R.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (24-JAN-1999) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
2 (bases 1 to 129547)
Ecker,J.R.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (23-APR-1999) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 129547)
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C.,
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Tortum,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
4 (bases 1 to 129547)
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C.,
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Tortum,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (22-JUL-2000) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
5 (bases 1 to 129547)

AUTHORS Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C.,
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Tortum,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Apr 23, 1999 this sequence version replaced gi:4416544.
FEATURES
Source
CDS
1..129547
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(join<1..290,372..426))
/note="similar to receptor serine/threonine kinase-like
protein embjCA16797.1"
/codon_start=1
/evidence=not-experimental
/product="F10A5.1"
/protein_id="AAF87135.1"
/db_xref="GI:9369386"
/translation="MALPLPLFLIFSHLYSVGRSTFTIWNKCEYTWPGILSNAG
VPELPPTGFLVQKGEERTISAPVSWGFRFWRQCSPTDDEKFCPLTGDCSGTLECS
GSGATPPATLAEF"
join(1927..2064,2190..2255,2470..2643,2856..2958,
3056..3416,3520..3885,3968..4046,4126..4335)
/note="similar to L-ascorbate oxidase homolog precursor
sp|Q00624|ASO_BRANA"
/codon_start=1
/evidence=not-experimental
/product="F10A5.2"
/protein_id="AAF87105.1"
/db_xref="GI:9369356"
/translation="MRHVFVEVLIVSLVLELSYAPAPISSTVOMVYSQRFILGN
KOVIVINDMPGPIANTANDIIVYKQITSTFFPLPLQKAGGIGAIRIYPPELV
PVPPKDEEDYDILIGWFTLUDHTVMASLDAGHSIPNPGILFENGREPTEFAEP
GKTYRLKISVNGKLTCLNFRIOHDMLEVEGYVOKRVYSSDIHVGOYSILVTA
KTDPEVGYRSYVIFATRFVDSYLGIALIRGYSPDPYGGCLAPALQDFGSVVO
ALSIKMDLVGAARSNPOGSHYGRIVFTIILHNDVMSKGLRTYIGSVVPE
TPKLVDHPLNDITIPGMPVYPSNKTPLGTSDVTHYKDFTHIVFQNDIGESY
HIDGYNFVYVYSGFGAMSKKACNYIVDVASSTVOVYRYSTAILIADNOMMNV
RSKARQWYLGQELVMYKCEGEDSTIPVROENPVGWIKCGKR"
join(5713..6109,6903..7172,7261..7937)
/note="putative tubulin beta-4 chain sp|P50262|TBB4_PORPU;
similar to ESTs dbj|AV554594.1, dbj|AV536568.1,
dbj|AV543492.1, dbj|AV542234.1, and gb|W43629.1"
/codon_start=1
/evidence=not-experimental
/product="F10A5.3"
/protein_id="AAF87106.1"
/db_xref="GI:9369357"
/translation="MREILHVGQCGCNOIGSKFWEYICDEHGVPTGRYNGSDADQ
LERINIVYNEASGRVYRAVLMDLEPGTMDSTRSGPYQIFPPDVGSGAGANNM
AKGHTEGAEILADVIVRKAEKNCDCLOGRVCISGCTSGMGTILIKREYD
PDMKMTFVSFSPKSVDPVPEYKNTLSYHOLVENADGEMVINDKALIDICRTKL
STPSFEDDLHLISATMSGVYCSLRFPGQLSDRLKLAIVLIPPRLHFVWGVAPLTS
RGSQOYISLTPPELTIQOMDKMMCAADPRHGRVLTASMEGKSTKEVDQILNV
QKNNSYFEGEMEPNNYKSVCDLPTGIRKASTFFVGNSTSIQMFRRVSGFAMFR
KAFILWYTGEMMEFTFAESNMNLVSPYQYODATADEDEYDEEEFOYVES"
complement(join(8344..8471,8704..8782,8967..9293))
/note="hypothetical protein"
/codon_start=1
/evidence=not-experimental
/product="F10A5.4"
/protein_id="AAF87134.1"
/db_xref="GI:9369385"
/translation="MSKKKTKOSLSMEDRDKGKKRDPFISIDYEISILLOLQSH
VIFSSDPVLHMGKRTKRSSSSSHLHPPEVYKTSIPCMADVGDMSKSSSSSCTIGEA
KTNQIERSSEFFSPRGVHCYLFIFCEPICIKSESGFORVBNKVOVAQAQRRLGL

```

CDS
      join(10894..11116,11285..11484,11708..11848,11931..12002,
      12074..12199,12327..12375)
      /note="unknown protein; similar to EST dbj|AV559963.1"
      /codon_start=1
      /evidence=not_experimental
      /product="F10A5.5"
      /protein_id="AAE87107.1"
      /db_xref="GI:9369358"
      /translation="MKAATRPVAVTTWVRQPPKKGFLGVYSANTALVLYIYHD
      HDNLFAVAEAVHSLGISLVLYLKTKERTCAGLSLKSOELTALFLAVRLYCSFWEEDJ
      HTLDSATLVTLTVITVMIRFLKASMDKDNFAIYVVPVILHPSTHHDI
      INKISWAFVYLAASVLPOLRYMONTKPTTAHVVALGIAREFSCAHWVLYQVLDTRG
      RLITALGCGFWPVLVLSLSEIVOTFEIADFCYIYVYSLMGQQLVRLPSGVV"
      join(17031..17238,17321..17502)
      /note="similar to GAST-like gene product gb|AAE97006.1;
      similar to ESTs dbj|AV536142.1, and dbj|AV566951.1"
      /codon_start=1
      /evidence=not_experimental
      /product="F10A5.6"
      /protein_id="AAE87108.1"
      /db_xref="GI:9369359"
      /translation="MAISKALIASLISLVLYLVQADVRLPHHKIILINILIMESK
      TYPSHVTTFLEQNSOKNGYARKIDGSAVCARCLSRPRLCHRCAGTCYRCNCV
      PCGTGNYDKCCQYASLTTHGRRKCP"
      complement(18355..18426)
      /product="tRNA-Thr"
      /note="codon recognized: ACG: putative tRNA-Thr"
      /evidence=not_experimental
      join(20514..20675,20808..20865,21251..21363,21464..21753,
      21840..21900,21989..22025,22094..22157,22244..22643,
      22721..22852,22938..23324)
      /note="hypothetical protein"
      /codon_start=1
      /evidence=not_experimental
      /product="F10A5.7"
      /protein_id="AAE87109.1"
      /db_xref="GI:9369360"
      /translation="MDLGSFSTKSKKMKLKLGFADDEKLKQEMSYDLPSPDLENEY
      SKKSKPKKLNKDTNGVDHASVPRKLSAMKRNLESPVLSKLSVSRKLNKEGTGS
      VSKKSEEMAKAVTESMMISDKDEVAETLVGLACMFETTSGLCKTSEKRTSY
      VDSILVEDDYTTESLTLPVSVLSAKTKQIDMPLQSOQKQNTGKWDFTDLKQ
      SSSKVVDAHARVNETVTTNSGLAMPGLSTTVHGAOVLSKSPSTKLPWNGOAV
      SPNTASLNCDELPVQPRKLRKASHIYISRLIKVQTSLSIATINONQRSSEMS
      ERLPLDITINDEKAIIVSPKRYONPHLDMHRAHNPQVDEMTKLALETYCPHS
      OKGSPDLISAGAAOSHPLPNSPROYOISAAVNSOLSPATSHOOLSPYASRQ
      TATNANOQOQLOLKLMAAOYRPTNGNTMQSNQYSRPNLSLNTSTIDQPLQVSSSR
      YNHVSOOHRILMTAAAMSMSHOHNNPSRAVMNRQEHFPLIEDTRTPLQLLCNEO
      S"
      join(23997..24116,24212..24667)
      /note="hypothetical protein"
      /codon_start=1
      /evidence=not_experimental
      /product="F10A5.8"
      /protein_id="AAE87110.1"
      /db_xref="GI:9369361"
      /translation="METTKRAEIDTTPAFTVKEAVALEGERVALASOVYSHLKMHE
      KWEDPSGIKEIOTRYDILKRAKESQIOMRNSLSLKEELERTQOELQOKLRVDGVE
      TKDEYFKTKFEVLVPRVDEPIRSRLBSMSKRYVKFANPIGNNSVYLERHPSM
      KKKEKTKDKKRSKSLPLFVGIFGIFSKKRVLO"
      complement(25054..25338)
      /note="hypothetical protein"
      /codon_start=1
      /evidence=not_experimental
      /product="F10A5.9"
      /protein_id="AAE87113.1"
      /db_xref="GI:9369384"
      /translation="MKLFHKLGIIMRLIDFTLSSSRQKNNRSRSGDGERPEPKT
      SCSSSYSHVHYEAIADIEFPNKSAMSCDEHFAVLNNIRDCYV"
      complement(join(28187..28890,29737..29834,29923..30509))
      /note="unknown protein; similar to ESTs gb|BE038339.1,
      gb|R30294.1, dbj|AV545596.1, and gb|AT99244.2.1"
      /codon_start=1

```

```

      /evidence=not_experimental
      /product="F10A5.10"
      /protein_id="AAE87132.1"
      /db_xref="GI:9369383"
      /translation="MALLTFPAEAEAPQKHKPTSSRKRRKRDNPDTOTOKHKPKPK
      KRAVPKQPSMDDIKNLITCKQIEGSRVDPKSNQSGSMNTNLSPLKLGSSCSI

Query Match      100.0%; Score 15; DB 8; Length 129547;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAAC 15
Db      113975 AGCCATAACATGAAC 113989

RESULT 13
AC016036/c
LOCUS
DEFINITION
AC016036
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC016036
Homo sapiens clone RP11-23K12, LOW-PASS SEQUENCE SAMPLING.
AC016036
AC016036.3 GI:9128702
HTG: HTGS_PHASE0.
Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
1 (bases 1 to 145520)
Homo sapiens, clone RP11-23K12
Unpublished
2 (bases 1 to 145520)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckery, R., Boguslavsky, L., Bonkhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearliano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Gallagan, J., Gardyna, S., Grant, G., Hagos, B., Healdorf, A., Horton, L.,
Holland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lien, C., Locke, K., Macdonald, P., Marquis, N.,
McKean, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefayeh, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (19-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:7108033.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4386
Center clone name: 23_K_112

* NOTE: This record contains 159 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

1	715: contig of 715 bp in length	30894	30993: gap of 100 bp	30894	30993: gap of 100 bp
716	815: gap of 100 bp	30994	31729: contig of 736 bp in length	30994	31729: contig of 736 bp in length
816	1545: contig of 730 bp in length	31730	31829: gap of 100 bp	31730	31829: gap of 100 bp
1546	1645: gap of 100 bp	31830	32540: contig of 711 bp in length	31830	32540: contig of 711 bp in length
2358	2457: contig of 712 bp in length	32541	32640: gap of 100 bp	32541	32640: gap of 100 bp
2458	3201: contig of 744 bp in length	32641	33372: contig of 732 bp in length	32641	33372: contig of 732 bp in length
3202	3301: gap of 100 bp	33373	33472: gap of 100 bp	33373	33472: gap of 100 bp
4042	4041: contig of 740 bp in length	33473	34203: contig of 731 bp in length	33473	34203: contig of 731 bp in length
4142	4900: contig of 759 bp in length	34204	34303: gap of 100 bp	34204	34303: gap of 100 bp
4901	5000: gap of 100 bp	35039	35138: contig of 735 bp in length	35039	35138: contig of 735 bp in length
5001	5733: contig of 733 bp in length	35139	35875: contig of 737 bp in length	35139	35875: contig of 737 bp in length
5734	5833: gap of 100 bp	35876	35975: gap of 100 bp	35876	35975: gap of 100 bp
5834	6556: contig of 723 bp in length	35976	36711: contig of 736 bp in length	35976	36711: contig of 736 bp in length
6557	7377: contig of 721 bp in length	36712	36811: gap of 100 bp	36712	36811: gap of 100 bp
7378	7477: gap of 100 bp	36812	37547: contig of 736 bp in length	36812	37547: contig of 736 bp in length
7478	8204: contig of 727 bp in length	37548	37647: gap of 100 bp	37548	37647: gap of 100 bp
8205	8304: gap of 100 bp	37648	38408: contig of 761 bp in length	37648	38408: contig of 761 bp in length
9030	9129: gap of 100 bp	38409	38508: gap of 100 bp	38409	38508: gap of 100 bp
9130	9867: contig of 738 bp in length	38509	39244: contig of 736 bp in length	38509	39244: contig of 736 bp in length
9868	9967: gap of 100 bp	39245	39344: gap of 100 bp	39245	39344: gap of 100 bp
9968	10712: contig of 745 bp in length	39345	40060: contig of 716 bp in length	39345	40060: contig of 716 bp in length
10713	10812: gap of 100 bp	40061	40160: gap of 100 bp	40061	40160: gap of 100 bp
10813	11550: contig of 738 bp in length	40161	40893: contig of 733 bp in length	40161	40893: contig of 733 bp in length
11551	11650: gap of 100 bp	40894	40993: gap of 100 bp	40894	40993: gap of 100 bp
11651	12407: contig of 757 bp in length	40994	41721: contig of 728 bp in length	40994	41721: contig of 728 bp in length
12408	12507: gap of 100 bp	41722	41821: gap of 100 bp	41722	41821: gap of 100 bp
12508	13264: contig of 757 bp in length	41822	42570: contig of 749 bp in length	41822	42570: contig of 749 bp in length
13265	13364: gap of 100 bp	42571	42670: gap of 100 bp	42571	42670: gap of 100 bp
13365	14107: contig of 743 bp in length	42671	43409: contig of 739 bp in length	42671	43409: contig of 739 bp in length
14108	14207: gap of 100 bp	43410	43509: gap of 100 bp	43410	43509: gap of 100 bp
14208	14932: contig of 725 bp in length	43510	44264: contig of 755 bp in length	43510	44264: contig of 755 bp in length
14933	15032: gap of 100 bp	44265	44364: gap of 100 bp	44265	44364: gap of 100 bp
15033	15748: contig of 717 bp in length	44365	45101: contig of 737 bp in length	44365	45101: contig of 737 bp in length
15750	15849: gap of 100 bp	45102	45201: gap of 100 bp	45102	45201: gap of 100 bp
15850	16589: contig of 740 bp in length	45202	45947: contig of 746 bp in length	45202	45947: contig of 746 bp in length
16590	16689: gap of 100 bp	45948	46047: gap of 100 bp	45948	46047: gap of 100 bp
16690	17435: contig of 746 bp in length	46048	46791: contig of 744 bp in length	46048	46791: contig of 744 bp in length
17436	17535: gap of 100 bp	46792	46891: gap of 100 bp	46792	46891: gap of 100 bp
17536	18279: contig of 744 bp in length	46892	47619: contig of 728 bp in length	46892	47619: contig of 728 bp in length
18280	18379: gap of 100 bp	47620	47719: gap of 100 bp	47620	47719: gap of 100 bp
18380	19125: contig of 746 bp in length	47720	48444: contig of 725 bp in length	47720	48444: contig of 725 bp in length
19126	19225: gap of 100 bp	48445	48544: gap of 100 bp	48445	48544: gap of 100 bp
19226	19984: contig of 759 bp in length	48545	49273: contig of 729 bp in length	48545	49273: contig of 729 bp in length
19985	20084: gap of 100 bp	49274	49373: gap of 100 bp	49274	49373: gap of 100 bp
20085	20826: contig of 742 bp in length	49374	50076: contig of 703 bp in length	493	

Query Match 100.0%; Score 15; DB 2; Length 145520;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAC 15
 DB 125440 AGCCATACATGAC 125426

RESULT 14
 AC022311 154084 bp DNA linear PRI 09-MAY-2001
 LOCUS AC022311
 DEFINITION Homo sapiens BAC clone RP11-146M24 from 2, complete sequence.
 AC022311
 AC022311.5 GI:11120940
 VERSION
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE
 AUTHORS 2 (bases 1 to 154084)
 TITLE Watkins, C., Abbott, A. and Fleming, A.
 JOURNAL The sequence of Homo sapiens BAC clone RP11-146M24
 REFERENCE 3 (bases 1 to 154084)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (29-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA.

REFERENCE
 AUTHORS 4 (bases 1 to 154084)
 TITLE Waterston, R.H.
 JOURNAL Direct Submission
 REFERENCE 5 (bases 1 to 154084)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Nov 8, 2000 this sequence version replaced gi:7630864.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0146M24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis

MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osonegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-457A20; the clone sequenced to the right is RP11-36C8. Actual start of this clone is at base position 1 of RP11-146M24; actual end is at base position 154084 of RP11-146M24.

There are polymorphic base pair differences in the overlap between the clone RP11-146M24 and RP11-36C8.

FEATURES

source

1..154084
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-146M24"
 /clone_1lb="RPCI-11"
 1..1170
 /rpl_family="L1"
 143..213
 /rpl_family="L1"
 1583..1629
 /rpl_family="ERV1"
 3328..3532
 /rpl_family="MIR"
 3606..4024
 /rpl_family="MALR"
 5253..5307
 /rpl_family="MIR"
 5521..5988
 /rpl_family="MALR"
 8746..8906
 /rpl_family="MALR"
 8953..9017
 /rpl_family="ERV1"
 9095..9144
 /rpl_family="ERV1"
 9662..9773
 /rpl_family="MER1_type"
 10625..10916
 /rpl_family="L1"
 11095..11344
 /rpl_family="L1"
 11359..11662
 /rpl_family="MALR"
 11663..11719
 /rpl_family="MALR"
 11735..12351
 /rpl_family="MALR"
 12352..12520
 /rpl_family="MALR"
 12521..12885
 /rpl_family="MALR"
 12886..12959
 /rpl_family="MALR"
 13020..13216
 /rpl_family="MALR"
 13221..13351
 /rpl_family="MALR"
 13353..13692
 /rpl_family="MALR"

```

repeat_region      14168..14517
                    /rpl_family="MALR"
repeat_region      14531..14747
                    /rpl_family="L1"
repeat_region      14903..16651
                    /rpl_family="L1"
repeat_region      16655..16914
                    /rpl_family="L1"
repeat_region      16952..17403
                    /rpl_family="L1"
repeat_region      17415..17609
                    /rpl_family="L1"
repeat_region      18536..18614
                    /rpl_family="L1"
repeat_region      20111..20186
                    /rpl_family="MIR"
repeat_region      20237..20516
                    /rpl_family="MIR"
repeat_region      20867..21389
                    /rpl_family="L2"
repeat_region      23662..23850
                    /rpl_family="MERL_type"
repeat_region      23864..23960
                    /rpl_family="L1"
repeat_region      23984..24136
                    /rpl_family="L1"
repeat_region      25154..25483
                    /rpl_family="MALR"
repeat_region      25563..25610
                    /rpl_family="L1"
repeat_region      26366..26464
                    /rpl_family="MERL_type"
repeat_region      26631..26829
                    /rpl_family="MIR"
repeat_region      28319..28616
                    /rpl_family="L1"
repeat_region      29095..29373
                    /rpl_family="L1"
repeat_region      29391..29660
                    /rpl_family="L1"
repeat_region      30983..31066
                    /rpl_family="MIR"
repeat_region      31668..32325
                    /rpl_family="L2"
repeat_region      32326..32638
                    /rpl_family="L1"
repeat_region      32639..32812
                    /rpl_family="L1"
repeat_region      32830..33552
                    /rpl_family="L1"
repeat_region      33597..33964
                    /rpl_family="L1"
repeat_region      34022..34096
                    /rpl_family="L1"
repeat_region      34097..34392
                    /rpl_family="L1"
repeat_region      34393..35687
                    /rpl_family="L1"
repeat_region      35818..36495
                    /rpl_family="L1"
repeat_region      36346..36601
                    /rpl_family="L1"
repeat_region      36654..38091
                    /rpl_family="L1"
repeat_region      38084..39798
                    /rpl_family="L1"

```

Query Match 100.0%; Score 15; DB 9; Length 154084;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15
 |||||
 DB 91395 AGCCATACATGAC 91409

RESULT 15
 AC130009/c
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-324C5, *** SEQUENCING IN PROGRESS
 ***, 55 unordered pieces.
 AC130009 158861 bp DNA linear HTG 08-AUG-2002
 AC130009
 AC130009.1 GI:22138172
 HTG: HTGS.PHASE1.
 SOURCE
 Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus

REFERENCE
 AUTHORS
 1 (bases 1 to 158861)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Adnan,F.R., Allen,C.,
 Alshrocks,S.L., Amaralungue,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.F.,
 Buhaq,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
 Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.,
 Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,D., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorielli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C., Hollins,L.E.,
 Homs,J.F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Medford,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G.,
 Otaguene,N., Oyledo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 158861)
 Worley,K.C.
 Direct Submission
 Submitted (08-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KAPP
 Center clone name: CH230-324C5
 ----- Summary Statistics

```

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124544 bases at least Q40
Consensus quality: 128196 bases at least Q30
Consensus quality: 131011 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hqsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 2609: contig of 1307 bp in length
* 2610 2709: gap of unknown length
* 2710 3850: contig of 1141 bp in length
* 3851 3950: gap of unknown length
* 3951 4978: contig of 1028 bp in length
* 4979 5078: gap of unknown length
* 5079 6269: contig of 1191 bp in length
* 6270 6369: gap of unknown length
* 6370 7473: contig of 1104 bp in length
* 7474 7573: gap of unknown length
* 7574 8907: contig of 1334 bp in length
* 8908 9007: gap of unknown length
* 9009 10051: contig of 1044 bp in length
* 10052 10151: gap of unknown length
* 10152 11191: contig of 1040 bp in length
* 11192 11291: gap of unknown length
* 11292 12427: contig of 1136 bp in length
* 12428 12527: gap of unknown length
* 12528 13822: contig of 1195 bp in length
* 13823 15030: gap of unknown length
* 15031 15130: contig of 1208 bp in length
* 15131 17569: gap of unknown length
* 17570 17669: gap of unknown length
* 17670 19479: contig of 1810 bp in length
* 19480 19579: gap of unknown length
* 19580 20965: contig of 1386 bp in length
* 20966 22711: gap of unknown length
* 22712 22811: contig of 1646 bp in length
* 22812 24497: gap of unknown length
* 24498 24597: contig of 1686 bp in length
* 24598 26106: gap of unknown length
* 26107 26206: contig of 1509 bp in length
* 26207 27767: gap of unknown length
* 27768 27867: contig of 1561 bp in length
* 27868 28973: gap of unknown length
* 28974 29073: contig of 1106 bp in length
* 29074 30550: gap of unknown length
* 30551 30650: contig of 1477 bp in length
* 30651 32496: gap of unknown length
* 32497 32596: contig of 1846 bp in length
* 32597 33896: gap of unknown length
* 33897 33996: contig of 1300 bp in length
* 33997 35948: gap of unknown length
* 35949 36048: contig of 1952 bp in length
* 36049 37884: gap of unknown length
* 37885 37984: contig of 1836 bp in length
* 37985 39844: gap of unknown length
* 39845 39944: contig of 1860 bp in length
* 39945 41441: gap of unknown length
* 41442 41541: contig of 1497 bp in length
* 41542 42674: gap of unknown length
* 42675 42774: contig of 1133 bp in length
* 42775 42874: gap of unknown length

```

```

* 42775 44391: contig of 1617 bp in length
* 44392 44491: gap of unknown length
* 44492 46237: contig of 1746 bp in length
* 46238 46337: gap of unknown length
* 46338 47773: contig of 1436 bp in length
* 47774 47873: gap of unknown length
* 47874 51167: contig of 3294 bp in length
* 51168 51267: gap of unknown length
* 51268 53726: contig of 2459 bp in length
* 53727 53826: gap of unknown length
* 53827 56626: contig of 2800 bp in length
* 56627 56726: gap of unknown length
* 56727 60004: contig of 3278 bp in length
* 60005 60104: gap of unknown length
* 60105 63599: contig of 3495 bp in length
* 63600 63699: gap of unknown length
* 63700 66017: contig of 2318 bp in length
* 66018 66117: gap of unknown length
* 66118 68050: contig of 1933 bp in length
* 68051 68150: gap of unknown length
* 68151 72358: contig of 4208 bp in length
* 72359 72458: gap of unknown length
* 72459 74954: contig of 2496 bp in length
* 74955 75054: gap of unknown length
* 75055 78265: contig of 3211 bp in length
* 78266 78365: gap of unknown length
* 78366 81407: contig of 3042 bp in length
* 81408 81507: gap of unknown length
* 81508 84643: contig of 3136 bp in length
* 84644 84743: gap of unknown length
* 84744 88452: contig of 3709 bp in length
* 88453 88552: gap of unknown length
* 88553 93253: contig of 4701 bp in length
* 93254 93353: gap of unknown length
* 93354 98211: contig of 4858 bp in length
* 98212 98311: gap of unknown length
* 98312 102965: contig of 4654 bp in length
* 102966 103065: gap of unknown length
* 103066 107049: contig of 3984 bp in length
* 107050 107149: gap of unknown length
* 107150 113981: contig of 6832 bp in length
* 113982 114081: gap of unknown length
* 114082 119381: contig of 5300 bp in length
* 119382 119481: gap of unknown length
* 119482 125815: contig of 6334 bp in length
* 125816 125915: gap of unknown length
* 125916 132747: contig of 6832 bp in length
* 132748 132847: gap of unknown length
* 132848 141359: contig of 8492 bp in length
* 141360 141459: gap of unknown length
* 141460 148269: contig of 6830 bp in length
* 148270 148369: gap of unknown length
* 148370 158861: contig of 10492 bp in length.

```

```

FEATURES
    Source
        1..158861
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-324C5"

```

```

Query Match      100.0%  Score 15:  DB 2:  Length 158861:
Best Local Similarity 100.0%:  Pred. No. 1.9e+02:
Matches 15:  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
Oy      1 AGCCATACATGAC 15
Db      115471 AGCCATACATGAC 115457

```

Search completed: June 26, 2003, 03:47:37
 Job time : 430.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :
1: N.Geneseq_101002:*
2: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID22/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	AA32405	Ab5 variable heavy
2	14	93.3	2401	ABL22930	Drosophila melanog
3	14	93.3	5235	ABL08064	Drosophila melanog
4	14	93.3	10425	ABL13480	Drosophila melanog
5	14	93.3	10425	ABL19870	Drosophila melanog
6	13.4	89.3	306	AA556643	Eucalyptus grandis
7	13.4	89.3	348	AA55981	Eucalyptus grandis
8	13.4	89.3	428	AA554150	Arabidopsis thaliana
9	13.4	89.3	521	AA43836	Chicken secreted e

C	10	13.4	89.3	556	22	ABA62712	Human foetal liver
C	11	13.4	89.3	556	22	ABA30013	Probe #8479 for ge
C	12	13.4	89.3	556	22	AAK11091	Human brain expres
C	13	13.4	89.3	556	22	AAK36915	Human bone marrow
C	14	13.4	89.3	556	22	AA117764	Probe #7697 for ge
C	15	13.4	89.3	556	22	AA142716	Probe #11402 used
C	16	13.4	89.3	556	24	AB510923	Human genome-dariv
C	17	13.4	89.3	597	21	AC344482	Arabidopsis thaliana
C	18	13.4	89.3	731	11	AA05977	Gas vesicle proteol
C	19	13.4	89.3	1001	24	ABK51547	DNA upstream of Ha
C	20	13.4	89.3	1047	20	AA331559	Bacillus species p
C	21	13.4	89.3	1047	22	AA686597	DNA encoding a Bac
C	22	13.4	89.3	1183	23	AA559699	Propionibacterium
C	23	13.4	89.3	1416	22	AAK70918	Human immune/haema
C	24	13.4	89.3	1517	21	AA65319	Human secreted pro
C	25	13.4	89.3	1518	21	AA65348	Human secreted pro
C	26	13.4	89.3	1612	7	AA60392	Sequence encoding
C	27	13.4	89.3	1685	24	ABN83054	Human development
C	28	13.4	89.3	1687	18	AA784244	Human ICAM-4 gene
C	29	13.4	89.3	1687	19	AAV11684	Human ICAM-4 upstr
C	30	13.4	89.3	1719	21	AA172859	Chicken GW75 CDNA.
C	31	13.4	89.3	2066	22	AAE27219	Human neovasculari
C	32	13.4	89.3	2181	21	AA299576	CDNA encoding a hu
C	33	13.4	89.3	2340	20	AA80866	DNA encoding human
C	34	13.4	89.3	2354	21	AA550348	Human secreted pro
C	35	13.4	89.3	2393	21	AA46135	CDNA encoding an a
C	36	13.4	89.3	2430	21	AA172855	CDNA clone AR2, AT
C	37	13.4	89.3	2457	22	AA509680	Oat CDNA encoding
C	38	13.4	89.3	3355	20	AA36362	Human TIE ligand N
C	39	13.4	89.3	3355	21	AA77529	Human PRO188 CDNA
C	40	13.4	89.3	3355	22	AA521382	Human CDNA sequenc
C	41	13.4	89.3	3355	22	AA67379	Human angiogenesis
C	42	13.4	89.3	3355	24	AA031564	Human TIE ligand N
C	43	13.4	89.3	6306	22	ABA15649	Human nervous syst
C	44	13.4	89.3	6695	22	ABA15648	Human nervous syst
C	45	13.4	89.3	6697	22	ABA15650	Human nervous syst

ALIGNMENTS

RESULT 1	AA32405	standard; DNA: 15 BP.
ID	AA32405	
AC	AA32405:	
XX		
XX		
DT	17-JUN-1999	(first entry)
XX		
DE	Ab5 variable heavy (VH) chain CDR1 encoding DNA.	
KW	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;	
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;	
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;	
KW	myelodysplasia; myelotoxic chemotherapy; leukemia; tumour; MDSK; CDR;	
KW	neuromuscular; muscular dystrophy; complementarity determining region;	
KW	variable heavy chain; variable light chain; VH; VL; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	W09910494-A2.	
XX		
PD	04-MAR-1999.	
XX		
PF	21-AUG-1998;	98WO-US17364.
XX		
PR	25-AUG-1997;	97US-0918148.
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;	
XX		
DR	WPI: 1999-204666/17.	

DR P-PSDB; AAY06704.

XX New thrombopoietin receptor agonist antibodies - useful for
PT treating immunological or hematological disorders
XX

PS Claim 10; Page 80; 86pp; English.

XX The invention relates to an agonist antibody (Ab) which binds to a
CC thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
CC be used in the same way and for the same indications as thrombopoietin
CC (TPO). They can stimulate proliferation, differentiation or growth of
CC megakaryocytes. They may also be able to stimulate megakaryocytes to
CC increase platelet production. They can be used for treating
CC immunological or hematopoietic disorders, especially thrombocytopenia.
CC Thrombocytopenia - associated bone marrow hypoplasia (e.g. aplastic anemia
CC following chemotherapy or bone marrow transplant) may be effectively
CC treated with the antibody compounds as well as disorders such as
CC disseminated intravascular coagulation (DIC), immune thrombocytopenia
CC (HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
CC congenital thrombocytopenia, thrombotic thrombocytopenia and
CC myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
CC treatment of solid tumours or leukaemia, myeloblastic chemotherapy for
CC atolous or allogeneic bone marrow transplant, myelodysplasia,
CC idiopathic aplastic anemia, congenital thrombocytopenia, and immune
CC thrombocytopenia. The antibodies which bind to the TPO-R can be
CC used for improving neuromuscular function in a patient, e.g. in muscular
CC dystrophy. The products can also be used for detection and diagnosis. The
CC antibodies have a longer half-life than the natural ligand for the TPO-R.
CC Sequences AAX32387-X32413 represent DNA fragments encoding the CDR1,
CC CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
CC (VL) chains of antibodies Ab1 to Ab6.
XX

SO Sequence 15 BP; 7 A; 4 C; 2 G; 2 T; 0 other;

Query Match 100.0%; Score 15; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15
1 | | | | | | | | | | | | | | | | | |
Db 1 AGCCATACATGAC 15

RESULT 2
AB122930

ID ABL22930 standard; DNA; 2401 BP.

AC ABL22930;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 20263.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

PS Claim 1; SEQ ID NO 20263; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SO Sequence 2401 BP; 687 A; 464 C; 603 G; 647 T; 0 other;

Query Match 93.38; Score 14; DB 23; Length 2401;
Best Local Similarity 100.0%; Pred. No. 4,4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCATACATGAC 15
2 | | | | | | | | | | | | | | | | | |
Db 672 GCCATACATGAC 685

RESULT 3
AB108064/C

ID ABL08064 standard; cDNA; 5235 BP.

XX ABL08064;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 18674.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB63961.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

PS Claim 1; SEQ ID NO 18674; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 5235 BP; 1265 A; 1209 C; 1279 G; 1482 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 5235;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATACATGAA 14
DB 960 AGCCATACATGAA 947
|||||
ABLI3480 standard; cDNA; 10425 BP.
AC ABLI3480;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34922.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR P-PSDB; ABB69377.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 34922; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABB57737-ABB72072).
XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 10425 BP; 2780 A; 2460 C; 2408 G; 2777 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 10425;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCATACATGAC 15
|||||

DB 4051 GCCATACATGAC 4064

RESULT 5
ABLI9870
ID ABLI9870 standard; DNA; 10426 BP.
XX
XX ABLI9870;
AC
AC ABLI9870;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11083.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 11083; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABB57737-ABB72072).
XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 10426 BP; 2780 A; 2460 C; 2409 G; 2777 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 10426;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCATACATGAC 15
DB 4051 GCCATACATGAC 4064
|||||
AAC56643/C
ID AAC56643 standard; DNA; 306 BP.
XX
XX AAC56643;
AC
AC AAC56643;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #514.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

KM poplar: sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KM basic helix-loop-helix zipper; homeotic; homeobox; MADS;
 KM homeobox zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KM type 2 Cys2His2; CCAAT box element; MYB; ss.

OS Eucalyptus grandis.

PN WO200053724-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-US06112.

PR 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

PI WPI; 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide

Claim 1; Page 482; 747pp; English.

The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant, e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeobox/MADS, homeobox zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

SQ Sequence 306 BP; 97 A; 49 C; 80 G; 80 T; 0 other;

Query Match 89.3%; Score 13.4; DB 21; Length 306;

Best Local Similarity 93.3%; Pred. No. 8.3e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15

DB 237 AGCCATGACATGAC 223

RESULT 7
 AAC55981/c
 ID AAC55981 standard; DNA; 348 BP.

AC AAC55981;

DT 25-JAN-2001 (first entry)

DE Eucalyptus grandis transcription factor DNA sequence #112.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;

KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;

KW homeobox zipper; LIM domain; AP2; ERBS; zinc finger domain;

KW type 2 Cys2His2; CCAAT box element; MYB; ss.

OS Eucalyptus grandis.

PN WO200053724-A2.

XX 14-SEP-2000.

PF 09-MAR-2000; 2000WO-US06112.

PR 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

PI WPI; 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide

Claim 1; Page 73; 747pp; English.

The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant, e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeobox/MADS, homeobox zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

SQ Sequence 348 BP; 106 A; 62 C; 87 G; 93 T; 0 other;

Query Match 89.3%; Score 13.4; DB 21; Length 348;

Best Local Similarity 93.3%; Pred. No. 8.3e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15

DB 237 AGCCATGACATGAC 223

RESULT 8
 AAC54150/c
 ID AAC54150 standard; DNA; 428 BP.

AC AAC54150;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 76865.

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136382.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 06-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.

```

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      89.3%; Score 13.4; DB 21; Length 428;
Best Local Similarity 93.3%; Pred. No. 8.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 AGCCATACATGAC 15
    |||
Db 419 AGCCATACATGAC 405

```

```

RESULT 9
AAA43836
ID AAA43836 standard; cDNA; 521 BP.
XX
AC AAA43836;

```

```

XX 21-AUG-2000 (first entry)
XX
DE Chicken secreted expressed sequence tag SEQ ID NO:411.
XX

```

Human; mouse; chicken; rat; secreted expressed sequence tag; SEST; expressed sequence tag; EST; probe; chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antisthmatic; vulnery; antiparkinsonian; antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; autoimmune disorder; multiple sclerosis; allergic condition; insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Huntington's disease; stroke; Parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss.

XX Gallus sp.

OS WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

XX 15-OCT-1999; 98US-0104436.

```

XX (GEMV ) GENETICS INST INC.
PA
XX
XX Jacobs K, McCoy JM, Lavallee ER, Collins-Racie LA, Evans C;
PI Metberg D, Treacy W, Bowman MR;
XX
XX WPI; 2000-317938/27.
XX

```

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (SESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders -

Claim 1: Page 310-311; 803pp; English.

AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antisthmatic; vulnery; antitumor; osteoprotective; neuroprotective; nootropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention.

Sequence 521 BP; 177 A; 104 C; 125 G; 115 T; 0 other;

```

Query Match      89.3%; Score 13.4; DB 21; Length 521;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 AGCCATACATGAC 15
    |||
Db 387 AGCCATACATGAC 401

```

```

RESULT 10
ABA62712/c
ID ABA62712 standard; DNA; 556 BP.
XX
AC ABA62712;

```

```

XX 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #11017.
XX

```

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1: SEQ ID NO 11017; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAC 15
Db | | | | | | | | | | | | | | | | | |
34 AGCCATACATTAAC 20

RESULT 11
ABA30013/C
ID ABA30013 standard; DNA; 556 BP.
XX
XX ABA30013:
AC
XX 23-JAN-2002 (first entry)
DT
XX
XX Probe #8479 for gene expression analysis in human heart cell sample.
DE
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
XX Homo sapiens.
OS
XX W0200157274-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PE
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT

PT hearts -
XX
XX Claim 1: SEQ ID NO 8479; 530pp; English.
PS
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAC 15
Db | | | | | | | | | | | | | | | | | |
34 AGCCATACATTAAC 20

RESULT 12
AAK11091/C
ID AAK11091 standard; DNA; 556 BP.
XX
XX AAK11091:
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe SEQ ID NO: 11082.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
OS
XX W0200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PE
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
PT
XX
XX Example 4: SEQ ID NO: 11082; 650pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;

Best Local Similarity 93.3%; Pred. No. 8.5e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
Db 34 AGCCATACATGAAC 20
|||||

RESULT 13

AAK36915/C
ID AAK36915 standard; DNA: 556 BP.

XX AAK36915;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 11472.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX MO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 11472; 658bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples, which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukaemia and myeloma. The present sequence is one of

XX the probes of the invention.

XX Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;

Best Local Similarity 93.3%; Pred. No. 8.5e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
Db 34 AGCCATACATGAAC 20
|||||

RESULT 14
AA117764/C
ID AA117764 standard; DNA: 556 BP.

XX AA117764;

XX 12-OCT-2001 (first entry)

XX Probe #7697 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.

XX Homo sapiens.

XX MO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 7697; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENPs). The present sequence is one such probe. The SENPs are derived

XX from human HeLa cells. The SENPs can be used to produce a single exon

XX microarray, which can be used for measuring human gene expression in a

XX sample derived from human cervical epithelial cells. By measuring gene

XX expression, the probes are therefore useful in grading and/or staging

XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;

Best Local Similarity 93.3%; Pred. No. 8.5e+02;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
Db 34 AGCCATACATGAAC 20
|||||

AA142716/C
ID AA142716 standard; DNA: 556 BP.

XX AA142716;

XX 17-OCT-2001 (first entry)

XX Probe #11402 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX

KW genetic disorder: ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI: 2001-488897/53.
 DK
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 PT
 XX
 PS Claim 25; SEQ ID No 11402; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 XX
 SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other:
 Query Match 89.3%; Score 13.4; DB 22; Length 556;
 Best Local Similarity 93.3%; Pred. No. 8.5e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGCCATACATGAC 15
 |||||
 Db 34 AGCCATAACATAAAC 20

Search completed: June 26, 2003, 03:11:49
 Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0', Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	US-08-918-148-36	Sequence 36, Appl
2	13.4	89.3	1047	US-09-198-955A-7	Sequence 7, Appl
3	13.4	89.3	1047	US-09-694-531-7	Sequence 7, Appl
4	13.4	89.3	1687	US-08-656-984A-33	Sequence 33, Appl
5	13.4	89.3	3355	US-08-933-821-3	Sequence 3, Appl
6	13.4	89.3	3355	US-08-960-507-3	Sequence 3, Appl
7	13.4	89.3	3355	US-09-136-828-3	Sequence 3, Appl
8	13.4	89.3	3355	US-09-332-928A-3	Sequence 3, Appl
9	13.4	89.3	3355	US-09-136-801-3	Sequence 3, Appl
10	13.4	89.3	3355	US-09-332-929-3	Sequence 3, Appl
11	13.4	86.7	477	US-08-227-357-129	Sequence 129, Appl
12	13	86.7	3106	US-08-840-466A-21	Sequence 21, Appl
13	13	86.7	3106	US-09-696-188B-21	Sequence 21, Appl
14	13	86.7	3131	US-08-840-466A-20	Sequence 20, Appl
15	13	86.7	3131	US-09-696-188B-20	Sequence 20, Appl
16	13	86.7	43360	US-09-453-702B-206	Sequence 206, Appl
17	13	86.7	45325	US-09-453-702B-261	Sequence 261, Appl
18	12.4	82.7	66	US-08-009-265-50	Sequence 50, Appl
19	12.4	82.7	66	US-08-009-265-51	Sequence 51, Appl
20	12.4	82.7	213	US-08-009-265-29	Sequence 29, Appl
21	12.4	82.7	213	PCT-US94-06079-45	Sequence 45, Appl
22	12.4	82.7	323	US-08-638-931-64	Sequence 64, Appl
23	12.4	82.7	324	US-08-905-223-264	Sequence 264, Appl
24	12.4	82.7	431	US-08-887-534A-3	Sequence 4, Appl
25	12.4	82.7	485	US-09-177-325-4	Sequence 4, Appl
26	12.4	82.7	485	US-09-411-812A-4	Sequence 4, Appl
27	12.4	82.7	485	US-09-590-113-4	Sequence 4, Appl

C 28	12.4	82.7	639	4	US-09-328-111-180	Sequence 180, App
C 29	12.4	82.7	970	3	US-08-888-077A-28	Sequence 28, Appl
C 30	12.4	82.7	1278	2	US-08-909-965C-4	Sequence 4, Appl
C 31	12.4	82.7	1536	4	US-09-352-990-17	Sequence 17, Appl
C 32	12.4	82.7	3088	1	US-08-418-444A-1	Sequence 1, Appl
C 33	12.4	82.7	3937	3	US-08-586-165-8	Sequence 8, Appl
C 34	12.4	82.7	9207	3	US-08-388-353-800	Sequence 800, App
C 35	12.4	82.7	9468	1	US-08-325-547-10	Sequence 10, Appl
C 36	12.4	82.7	9468	1	US-08-976-259-14	Sequence 14, Appl
C 37	12.4	82.7	32768	4	US-08-961-527-71	Sequence 71, Appl
C 38	12.4	82.7	36651	4	US-09-738-894A-3	Sequence 3, Appl
C 39	12.4	82.7	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 40	12.4	82.7	4411529	4	US-09-103-840A-1	Sequence 1, Appl
C 41	12.4	82.7	4411529	4	US-08-887-534A-88	Sequence 88, Appl
C 42	12	80.0	441	4	US-09-134-001C-456	Sequence 456, App
C 43	12	80.0	558	4	US-09-381-862-6	Sequence 6, Appl
C 44	12	80.0	4973	4	US-08-961-527-45	Sequence 45, Appl
C 45	12	80.0	11384	4		

ALIGNMENTS

```
RESULT 1
US-08-918-148-36
: Sequence 36, Application US/08918148A
: Patent No. 6342220
: GENERAL INFORMATION:
: APPLICANT: Adams, Camellia
: APPLICANT: W.
: APPLICANT: Carter, Paul J.
: APPLICANT: Fendly, Brian M.
: APPLICANT: Gurney, Austin L.
: TITLE OF INVENTION: Agonist Antibodies
: FILE REFERENCE: P0979
: CURRENT APPLICATION NUMBER: US/08/918,148A
: CURRENT FILING DATE: 1997-08-25
: NUMBER OF SEQ ID NOS: 79
: SEQ ID NO 36
: LENGTH: 15
: TYPE: DNA
: ORGANISM: artificial
: FEATURE:
: NAME/KEY: 12D5scfv VH CDR1
: LOCATION: 1-15
: OTHER INFORMATION:
US-08-918-148-36

Query Match      100.0%; Score 15; DB 4; Length 15;
Best local Similarity 100.0%; Pred. No. 5.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY      1 AGCCATACATGAAC 15
Db      1 AGCCATACATGAAC 15

RESULT 2
US-09-198-955A-7/C
: Sequence 7, Application US/09198955A
: Patent No. 6187580
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schuelein, Martin
: APPLICANT: Lange, Niels E.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Molter, Soren
: APPLICANT: Glad, Sanne O. S.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Schmoor, Kirk
: APPLICANT: Kongsbak, Lars
: TITLE OF INVENTION: No. 6187580el Pectate Lyases
: FILE REFERENCE: 5378.200-US
```

```

: CURRENT APPLICATION NUMBER: US/09/198,955A
: CURRENT FILING DATE: 1998-11-24
: PRIOR APPLICATION NUMBER: 1343/97
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 1344/97
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/067,249
: PRIOR FILING DATE: 1997-12-02
: PRIOR APPLICATION NUMBER: 60/067,240
: PRIOR FILING DATE: 1997-12-02
: PRIOR APPLICATION NUMBER: 09/073,684
: PRIOR FILING DATE: 1998-05-06
: PRIOR APPLICATION NUMBER: 09/184,217
: PRIOR FILING DATE: 1998-11-02
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1047
: TYPE: DNA
: ORGANISM: Bacillus sp.
US-09-198-955A-7

```

```

Query Match          89.3%; Score 13.4; DB 4; Length 1047;
Best Local Similarity 93.3%; Pred. No. 79;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 AGCCATACATGAC 15
        |||||
Db       95 AGCCATACATGAC 81

```

```

RESULT 3
US-09-694-531-7/c
: Sequence 7, Application US/09694531
: Patent No. 6368843
: GENERAL INFORMATION:
: APPLICANT: Andersen, Lene N.
: APPLICANT: Schuelein, Martin
: APPLICANT: Lange, Niels E.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Moller, Soren
: APPLICANT: Glad, Sanne O. S.
: APPLICANT: Kauppinen, Markus S.
: APPLICANT: Schmitt, Kirk
: TITLE OF INVENTION: No. 6368843el Pectate Lyases
: FILE REFERENCE: 5378.200-US
: CURRENT APPLICATION NUMBER: US/09/694,531
: CURRENT FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 09/198,955
: PRIOR FILING DATE: 1998-11-24
: PRIOR APPLICATION NUMBER: 1343/97
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 1344/97
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/067,249
: PRIOR FILING DATE: 1997-12-02
: PRIOR APPLICATION NUMBER: 60/067,240
: PRIOR FILING DATE: 1997-12-02
: PRIOR APPLICATION NUMBER: 09/073,684
: PRIOR FILING DATE: 1998-05-06
: PRIOR APPLICATION NUMBER: 09/184,217
: PRIOR FILING DATE: 1998-11-02
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1047
: TYPE: DNA
: ORGANISM: Bacillus sp.
US-09-694-531-7

```

```

Query Match          89.3%; Score 13.4; DB 4; Length 1047;
Best Local Similarity 93.3%; Pred. No. 79;

```

```

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 AGCCATACATGAC 15
        |||||
Db       95 AGCCATACATGAC 81

```

```

RESULT 4
US-08-656-984A-33/c
: Sequence 33, Application US/08656984A
: Patent No. 5753502
: GENERAL INFORMATION:
: APPLICANT: Gallatin, W. Michael
: APPLICANT: Kilgannon, Patrick D.
: TITLE OF INVENTION: ICAM-4 Materials and Methods
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 233 South Wacker Drive, 6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/656,984A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/827,689
: FILING DATE: 27-JAN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/889,724
: FILING DATE: 26-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/894,061
: FILING DATE: 05-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/009,266
: FILING DATE: 22-JAN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/102,852
: FILING DATE: 05-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/245,295
: FILING DATE: 18-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/485,604
: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: WILLIAMS, JR. JOSEPH A.
: REGISTRATION NUMBER: 38,659
: REFERENCE/DOCKET NUMBER: 27866/33321
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-474-6300
: TELEFAX: 312-474-0448
: TELE: 25-3856
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1687 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-656-984A-33

```

```

Query Match          89.3%; Score 13.4; DB 1; Length 1687;
Best Local Similarity 93.3%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```


OY 1 AGCCATAACATGAC 15
||||| |||||
DB 450 AGCCATAACATGAC 436

RESULT 5

US-08-933-821-3
Sequence 3, Application US/08933821
Patent No. 5972338
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Tie Ligands
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,821
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1130
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3355 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-933-821-3
Query Match 89.3%; Score 13.4; DB 2; Length 3355;
Best Local Similarity 93.3%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAC 15
||||| |||||
DB 815 AGCCATAACATGAC 829

RESULT 6

US-08-960-507-3
Sequence 3, Application US/08960507
Patent No. 6057435
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Tie Ligands
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,507
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1130P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3355 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-960-507-3

Query Match 89.3%; Score 13.4; DB 3; Length 3355;
Best Local Similarity 93.3%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAC 15
||||| |||||
DB 815 AGCCATAACATGAC 829

RESULT 7

US-09-136-828-3
Sequence 3, Application US/09136828
Patent No. 6350450
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Tie Ligands
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,828
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1130R1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3355 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-09-136-828-3

Query Match 89.3%; Score 13.4; DB 4; Length 3355;
Best Local Similarity 93.3%; Pred. No. 93;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
|||||

DB 815 AGCCGTACATGAAC 829

RESULT 8

US-09-332-928A-3

; Sequence 3, Application US/09332928A

; Patent No. 6368853

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; Gurney, Austin L.

; TITLE OF INVENTION: Tie ligands

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Winpatin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/332,928A

; FILING DATE: 14-Jun-1999

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/933,821

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1130

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-3216

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3355 base pairs

; TYPE: Nucleic Acid

; STRANDEDNESS: Single

; TOPOLOGY: Linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-332-928A-3

Query Match 89.3%; Score 13.4; DB 4; Length 3355;

Best Local Similarity 93.3%; Pred. No. 93;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
|||||

DB 815 AGCCGTACATGAAC 829

RESULT 9

US-09-136-801-3

; Sequence 3, Application US/09136801

; Patent No. 6413770

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; Gurney, Austin L.

; APPLICANT: Hillan, Kenneth

; APPLICANT: Botstein, David

; APPLICANT: Goddard, Audrey

; APPLICANT: Roy, Margaret

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Tumas, Daniel

; APPLICANT: Schwall, Ralph

; TITLE OF INVENTION: Tie Ligand Homologues

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Winpatin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,801

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1130P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-3216

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3355 base pairs

; TYPE: Nucleic Acid

; STRANDEDNESS: Single

; TOPOLOGY: Linear

US-09-136-801-3

Query Match 89.3%; Score 13.4; DB 4; Length 3355;

Best Local Similarity 93.3%; Pred. No. 93;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
|||||

DB 815 AGCCGTACATGAAC 829

RESULT 10

US-09-332-929-3

; Sequence 3, Application US/09332929

; Patent No. 6420542

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; Gurney, Austin L.

; TITLE OF INVENTION: Tie Ligands

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Winpatin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/332,929

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/933,821

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

```
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1130
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3355 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-09-332-929-3

Query Match      89.3%: Score 13.4: DB 4: Length 3355;
Best Local Similarity 93.3%: Pred. No. 93;
Matches 14: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
```

QY 1 AGCCATACATGAC 15
Db 815 AGCCGTACATGAC 829

```
RESULT 11
US-09-227-357-129
Sequence 129, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
FILE REFERENCE: P2010P1
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684
EARLIER FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
```

```
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 129
LENGTH: 477
TYPE: DNA
ORGANISM: Homo sapiens
US-09-227-357-129

Query Match      86.7%: Score 13: DB 4: Length 477;
Best Local Similarity 100.0%: Pred. No. 1,2e+02;
Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
```

QY 3 CCATACATGAC 15
Db 254 CCATACATGAC 266

```
RESULT 12
US-08-840-466A-21
Sequence 21, Application US/08840466A
Patent No. 6261561
GENERAL INFORMATION:
APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegau, Henderson, Farabow, Garrett &
Dunnet, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,466A
FILING DATE: 18-Apr-1997
CLASSIFICATION: <Unknown>
```

ATTORNEY/AGENT INFORMATION:
NAME: Boone, Laura S.
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04995.0029-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-08-840-466A-21

Query Match 86.7%; Score 13; DB 4; Length 3106;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGA 13
|||||
Db 60 AGCCATAACATGA 72

RESULT 13

US-09-696-188B-21
Sequence 21, Application US/09696188B
Patent No. 6406885

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
O'Brien, Allison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/696,188B
FILING DATE: 26-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/840,466
FILING DATE: 1997-04-18

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laura S.
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04995.0029-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-696-188B-21

Query Match 86.7%; Score 13; DB 4; Length 3106;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGA 13
|||||
Db 60 AGCCATAACATGA 72

RESULT 14

US-08-840-466A-20
Sequence 20, Application US/08840466A
Patent No. 6261561

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
O'Brien, Allison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,466A
FILING DATE: 18-Apr-1997
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laura S.
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04995.0029-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-08-840-466A-20

Query Match 86.7%; Score 13; DB 4; Length 3131;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGA 13
|||||
Db 197 AGCCATAACATGA 209

RESULT 15

US-09-696-188B-20
Sequence 20, Application US/09696188B

Patent No. 6406885
GENERAL INFORMATION:
APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Machtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/696,188B
FILING DATE: 26-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/840,466
FILING DATE: 1997-04-18
ATTORNEY/AGENT INFORMATION:
NAME: Boone, Laura L S
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04995.0029-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-696-188B-20

Query Match 86.7%; Score 13; DB 4; Length 3131;
Best local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATACATGA 13
|||||
Db 197 AGCCATACATGA 209

Search completed: June 26, 2003, 04:52:38
Job time : 31.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_MA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTOS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13.4	89.3	556	US-09-864-761-8479	Sequence 8479, App
2	13.4	89.3	655	US-09-897-306-5	Sequence 5, Appl
3	13.4	89.3	661	US-09-897-306-6	Sequence 6, Appl
4	13.4	89.3	776	US-09-910-943-412	Sequence 412, App
5	13.4	89.3	1047	US-10-072-152-7	Sequence 7, Appl
6	13.4	89.3	1391	US-09-938-842A-2733	Sequence 2733, App
7	13.4	89.3	1517	US-09-892-877-80	Sequence 80, Appl
8	13.4	89.3	1517	US-09-948-783-79	Sequence 79, Appl
9	13.4	89.3	1518	US-09-892-877-115	Sequence 115, App
10	13.4	89.3	1518	US-09-948-783-115	Sequence 115, App
11	13.4	89.3	1687	US-10-025-524-33	Sequence 33, Appl
12	13.4	89.3	2000	US-09-887-576-99	Sequence 99, Appl
13	13.4	89.3	2181	US-09-897-306-2	Sequence 2, Appl
14	13.4	89.3	2354	US-10-335-819-1	Sequence 1, Appl
15	13.4	89.3	3355	US-10-028-072-277	Sequence 277, App
16	13.4	89.3	3355	US-10-121-049-277	Sequence 277, App
17	13.4	89.3	3355	US-10-123-904-277	Sequence 277, App
18	13.4	89.3	3355	US-10-140-470-277	Sequence 277, App
19	13.4	89.3	3355	US-10-175-746-277	Sequence 277, App

20	13.4	89.3	3355	9	US-10-176-918-277	Sequence 277, App
21	13.4	89.3	3355	9	US-10-176-921-277	Sequence 277, App
22	13.4	89.3	3356	9	US-10-137-865-277	Sequence 277, App
23	13.4	89.3	3355	9	US-10-140-474-277	Sequence 277, App
24	13.4	89.3	3355	9	US-10-142-431-277	Sequence 277, App
25	13.4	89.3	3355	9	US-10-143-114-277	Sequence 277, App
26	13.4	89.3	3355	9	US-10-140-002-277	Sequence 277, App
27	13.4	89.3	3355	9	US-10-142-419-277	Sequence 277, App
28	13.4	89.3	3355	9	US-10-123-262-277	Sequence 277, App
29	13.4	89.3	3355	9	US-10-142-423-277	Sequence 277, App
30	13.4	89.3	3355	9	US-10-121-050-277	Sequence 277, App
31	13.4	89.3	3355	9	US-10-141-755-277	Sequence 277, App
32	13.4	89.3	3355	9	US-10-143-032-277	Sequence 277, App
33	13.4	89.3	3355	9	US-10-123-108-277	Sequence 277, App
34	13.4	89.3	3355	9	US-10-123-236-277	Sequence 277, App
35	13.4	89.3	3355	9	US-10-123-261-277	Sequence 277, App
36	13.4	89.3	3355	9	US-10-140-921-277	Sequence 277, App
37	13.4	89.3	3355	9	US-10-140-928-277	Sequence 277, App
38	13.4	89.3	3355	9	US-10-121-045-277	Sequence 277, App
39	13.4	89.3	3355	9	US-10-123-292-277	Sequence 277, App
40	13.4	89.3	3355	9	US-10-123-903-277	Sequence 277, App
41	13.4	89.3	3355	9	US-10-124-819-277	Sequence 277, App
42	13.4	89.3	3355	9	US-10-124-822-277	Sequence 277, App
43	13.4	89.3	3355	9	US-10-140-925-277	Sequence 277, App
44	13.4	89.3	3355	9	US-10-160-498-277	Sequence 277, App
45	13.4	89.3	3355	9	US-10-121-041-277	Sequence 277, App

ALIGNMENTS

RESULT 1
US-09-864-761-8479/c
; Sequence 8479, Application US/09864761
; Patient No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

```
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
;; SEQ ID NO: 8479
;; LENGTH: 556
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC018552.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
US-09-864-761-8479
```

```
Query Match      89.3%; Score 13.4; DB 10; Length 556;
Best Local Similarity 93.3%; Pred. No. 7e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 AGCCATACATGAAC 15
        |||||
DB      34 AGCCATACATGAAC 20
```

```
RESULT 2
US-09-897-306-5
;; Sequence 5, Application US/09897306
;; Patent No. US20020123054A1
;; GENERAL INFORMATION:
;; APPLICANT: Hillman, Jennifer L.
;; APPLICANT: Gorgone, Gina A.
;; APPLICANT: Patterson, Chandra
;; APPLICANT: Murty, Lynn E.
;; TITLE OF INVENTION: HUMAN ANGIOPOIETIN
;; FILE REFERENCE: PC-0048 CIP
;; CURRENT APPLICATION NUMBER: US/09/897,306
;; CURRENT FILING DATE: 2001-07-02
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: PERL Program
;; SEQ ID NO 5
;; LENGTH: 655
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incyte ID No. US20020123054A1 SBD07003F3
;; LOCATION: 532, 536, 555, 618, 630, 635, 641-642
;; OTHER INFORMATION: a, t, c, g, or other
US-09-897-306-5
```

```
Query Match      89.3%; Score 13.4; DB 10; Length 655;
Best Local Similarity 93.3%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 AGCCATACATGAAC 15
        |||||
DB      492 AGCCGTAACATGAAC 506
```

```
RESULT 3
US-09-897-306-6
```

```
;; Sequence 6, Application US/09897306
;; Patent No. US20020123054A1
;; GENERAL INFORMATION:
;; APPLICANT: Hillman, Jennifer L.
;; APPLICANT: Gorgone, Gina A.
;; APPLICANT: Patterson, Chandra
;; APPLICANT: Murty, Lynn E.
;; TITLE OF INVENTION: HUMAN ANGIOPOIETIN
;; FILE REFERENCE: PC-0048 CIP
;; CURRENT APPLICATION NUMBER: US/09/897,306
;; CURRENT FILING DATE: 2001-07-02
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: PERL Program
;; SEQ ID NO 6
;; LENGTH: 661
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incyte ID No. US20020123054A1 SBD05881F1
;; LOCATION: 475, 506, 510, 545, 547, 554, 581, 599-600, 603, 615, 633, 638, 658-659
;; OTHER INFORMATION: a, t, c, g, or other
US-09-897-306-6
```

```
Query Match      89.3%; Score 13.4; DB 10; Length 661;
Best Local Similarity 93.3%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 AGCCATACATGAAC 15
        |||||
DB      204 AGCCGTAACATGAAC 218
```

```
RESULT 4
US-09-910-943-412/C
;; Sequence 412, Application US/09910943
;; Patent No. US20020081610A1
;; GENERAL INFORMATION:
;; APPLICANT: Hemmati-Briandjou, Ali
;; APPLICANT: Altman, Curtis
;; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
;; FILE REFERENCE: 7529/IG14RUS1
;; CURRENT APPLICATION NUMBER: US/09/910,943
;; CURRENT FILING DATE: 2001-07-23
;; NUMBER OF SEQ ID NOS: 742
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 412
;; LENGTH: 776
;; TYPE: DNA
;; ORGANISM: Xenopus laevis
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(776)
;; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-412
```

```
Query Match      89.3%; Score 13.4; DB 10; Length 776;
Best Local Similarity 93.3%; Pred. No. 7.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 AGCCATACATGAAC 15
        |||||
DB      238 ATCCATACATGAAC 224
```

```
RESULT 5
US-10-072-152-7/C
;; Sequence 7, Application US/10072152
;; Patent No. US20020142438A1
;; GENERAL INFORMATION:
;; APPLICANT: Andersen, Lene N.
;; APPLICANT: Schulten, Martin
```



```
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Bacillus sp.
; US-10-072-152-7
```

```
Query Match      89.3%; Score 13.4; DB 12; Length 1047;
Best Local Similarity 93.3%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATACATGAAC 15
          |||||
Db      95 AGCCATACATGAAC 81
```

```
RESULT 6
US-09-938-842A-2733
; Sequence 2733, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2733
; LENGTH: 1391
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-2733
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1391;
Best Local Similarity 93.3%; Pred. No. 7.8e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATACATGAAC 15
          |||||
Db      281 AGCCATACATGAAC 295
```

```
RESULT 7
US-09-892-877-80/c
; Sequence 80, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 80
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1145)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-892-877-80
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1517;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATACATGAAC 15
          |||||
Db      756 AGCCAGACATGAAC 742
```

```
RESULT 8
US-09-948-783-79/c
; Sequence 79, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
```

;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,923
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,925
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,928
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,920
;; PRIOR FILING DATE: 1998-05-18
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 79
;; LENGTH: 1517
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (1145)
;; OTHER INFORMATION:
US-09-948-783-79

Query Match 89.3%; Score 13.4; DB 9; Length 1517;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15
||||| |||||||
DB 756 AGCCAGAACATGAC 742

RESULT 9
US-09-892-877-115/c
;; Sequence 115, Application US/09892877
;; Publication No. US20030077809A1
;; GENERAL INFORMATION:
;; APPLICANT: Ruben et. al.
;; TITLE OF INVENTION: 97 Human secreted proteins
;; FILE REFERENCE: P2028P1
;; CURRENT APPLICATION NUMBER: US/09/892,877
;; CURRENT FILING DATE: 2001-06-28
;; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
;; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
;; NUMBER OF SEQ ID NOS: 461
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 115
;; LENGTH: 1518
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (1146)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-892-877-115

Query Match 89.3%; Score 13.4; DB 9; Length 1518;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15
||||| |||||||
DB 757 AGCCAGAACATGAC 743

RESULT 10
US-09-948-783-115/c
;; Sequence 115, Application US/09948783
;; Publication No. US20030100051A1
;; GENERAL INFORMATION:
;; APPLICANT: Ruben et. al.
;; TITLE OF INVENTION: 97 Human secreted proteins
;; FILE REFERENCE: P2028P2
;; CURRENT APPLICATION NUMBER: US/09/948,783
;; CURRENT FILING DATE: 2001-09-10

;; PRIOR APPLICATION NUMBER: 60/231,846
;; PRIOR FILING DATE: 2000-09-11
;; PRIOR APPLICATION NUMBER: 09/892,877
;; PRIOR FILING DATE: 2001-06-28
;; PRIOR APPLICATION NUMBER: 09/437,658
;; PRIOR FILING DATE: 1999-11-10
;; PRIOR APPLICATION NUMBER: PCT/US99/09847
;; PRIOR FILING DATE: 1999-05-06
;; PRIOR APPLICATION NUMBER: 60/085,093
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,094
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,105
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,180
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,927
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,906
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,924
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,922
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,921
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,923
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,925
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,928
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,920
;; PRIOR FILING DATE: 1998-05-18
;; NUMBER OF SEQ ID NOS: 465
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 115
;; LENGTH: 1518
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (1146)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-115

Query Match 89.3%; Score 13.4; DB 9; Length 1518;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAC 15
||||| |||||||
DB 757 AGCCAGAACATGAC 743

RESULT 11
US-10-025-524-33/c
;; Sequence 33, Application US/10025524
;; Publication No. US20030068659A1
;; GENERAL INFORMATION:
;; APPLICANT: Gallatin, W. Michael
;; Kliganov, Patrick D.
;; TITLE OF INVENTION: ICAM-4 Materials and Methods
;; NUMBER OF SEQUENCES: 42
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 233 South Wacker Drive, 6300 Sears Tower
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/025,524
; FILING DATE: 18-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,689
; FILING DATE: 27-JAN-1992
; APPLICATION NUMBER: US 07/889,724
; FILING DATE: 26-MAY-1992
; APPLICATION NUMBER: US 07/894,061
; FILING DATE: 05-JUN-1992
; APPLICATION NUMBER: US 08/009,266
; FILING DATE: 22-JAN-1993
; APPLICATION NUMBER: US 08/102,852
; FILING DATE: 05-AUG-1993
; APPLICATION NUMBER: US 08/245,295
; FILING DATE: 18-MAY-1994
; APPLICATION NUMBER: US 08/485,604
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, JR. JOSEPH A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33321
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-10-025-524-33

Query Match      89.3%; Score 13.4; DB 9; Length 1687;
Best Local Similarity 93.3%; Pred. No. 8e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 AGCCATACATGAAC 15
Db      450 AGCCATACATGAAC 436

RESULT 12
US-09-887-576-99
; Sequence 99, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 99
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-99

Query Match      89.3%; Score 13.4; DB 10; Length 2000;
Best Local Similarity 93.3%; Pred. No. 8.1e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 AGCCATACATGAAC 15
Db      908 AGCCATACATGAAC 922

RESULT 13
US-09-897-306-2
; Sequence 2, Application US/09897306
; Patent No. US20020123054A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Patterson, Chandra
; APPLICANT: Morry, Lynn E.
; TITLE OF INVENTION: HUMAN ANGIOPOIETIN
; FILE REFERENCE: PC-0048 CIP
; CURRENT APPLICATION NUMBER: US/09/897,306
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 2181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020123054A1 2365223CB1
US-09-897-306-2

Query Match      89.3%; Score 13.4; DB 10; Length 2181;
Best Local Similarity 93.3%; Pred. No. 8.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 AGCCATACATGAAC 15
Db      552 AGCCATACATGAAC 566

RESULT 14
US-10-335-819-1
; Sequence 1, Application US/10335819
; Publication No. US20030104573A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Jeffers, Michael
; TITLE OF INVENTION: Nucleotide Sequences and Amino Acid Sequences of
; FILE REFERENCE: 15966-517 CIP1
; CURRENT APPLICATION NUMBER: US/10/335,819
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US/09/658,644
; PRIOR FILING DATE: 2001-09-09
; PRIOR APPLICATION NUMBER: U.S.S.N. 09/150,684
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2354
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (352)..(1824)
```

US-10-335-819-1

Query Match 89.3%; Score 13.4; DB 9;
 Best Local Similarity 93.3%; Pred. No. 8,3e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAAC 15
 |||||
 DB 718 AGCCGTACATGAC 732

RESULT 15

US-10-028-072-277
 ; Sequence 277, Application US/10028072
 ; Publication No. US20030004311A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

APPLICANT: Filvaroff, Ellen

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gettitsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang

TITLE OF INVENTION:

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/028.072

CURRENT FILING DATE: 2001-12-19

PRIOR APPLICATION NUMBER: 60/049911

PRIOR FILING DATE: 1997-06-18

PRIOR APPLICATION NUMBER: 60/056974

PRIOR FILING DATE: 1997-08-26

PRIOR APPLICATION NUMBER: 60/059113

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059115

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059117

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059122

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059184

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059352

PRIOR FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/059588

PRIOR FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/059836

PRIOR FILING DATE: 1997-09-24

PRIOR APPLICATION NUMBER: 60/062250

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/062285

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/062287

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/062814

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/062816

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063045

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063082

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/063127

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063327

PRIOR FILING DATE: 1997-10-27

PRIOR APPLICATION NUMBER: 60/063329

PRIOR FILING DATE: 1997-10-27

PRIOR APPLICATION NUMBER: 60/063550

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063561

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063704

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063733

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063735

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063738

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063755

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/064248

PRIOR FILING DATE: 1997-11-03

PRIOR APPLICATION NUMBER: 60/064809

PRIOR FILING DATE: 1997-11-07

PRIOR APPLICATION NUMBER: 60/065186

PRIOR FILING DATE: 1997-11-12

PRIOR APPLICATION NUMBER: 60/065846

PRIOR FILING DATE: 1997-11-17

PRIOR APPLICATION NUMBER: 60/066364

PRIOR FILING DATE: 1997-11-21

PRIOR APPLICATION NUMBER: 60/066453

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/066511

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/066770

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/069212

PRIOR FILING DATE: 1997-12-11

PRIOR APPLICATION NUMBER: 60/069278

PRIOR FILING DATE: 1997-12-11

PRIOR APPLICATION NUMBER: 60/069334

PRIOR FILING DATE: 1997-12-11

PRIOR APPLICATION NUMBER: 60/069694

PRIOR FILING DATE: 1997-12-16

PRIOR APPLICATION NUMBER: 60/072320

PRIOR FILING DATE: 1998-01-23

PRIOR APPLICATION NUMBER: 60/073612

PRIOR FILING DATE: 1998-02-04

PRIOR APPLICATION NUMBER: 60/074086

PRIOR FILING DATE: 1998-02-09

PRIOR APPLICATION NUMBER: 60/074092

PRIOR FILING DATE: 1998-02-09

PRIOR APPLICATION NUMBER: 60/077791

PRIOR FILING DATE: 1998-03-12

PRIOR APPLICATION NUMBER: 60/078910

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/079294

PRIOR FILING DATE: 1998-03-25

PRIOR APPLICATION NUMBER: 60/079663

PRIOR FILING DATE: 1998-02-27

PRIOR APPLICATION NUMBER: 60/079728

PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/080165

PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/081203

PRIOR FILING DATE: 1998-04-09

PRIOR APPLICATION NUMBER: 60/081229

PRIOR FILING DATE: 1998-04-09

PRIOR APPLICATION NUMBER: 60/081695

PRIOR FILING DATE: 1998-04-14

PRIOR APPLICATION NUMBER: 60/081817

PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/081818

; PRIOR FILING DATE: 1998-04-15
 ; PRIOR APPLICATION NUMBER: 60/082999
 ; PRIOR FILING DATE: 1998-04-24
 ; PRIOR APPLICATION NUMBER: 60/083322
 ; PRIOR FILING DATE: 1998-04-28
 ; PRIOR APPLICATION NUMBER: 60/083545
 ; PRIOR FILING DATE: 1998-04-29
 ; PRIOR APPLICATION NUMBER: 60/084600
 ; PRIOR FILING DATE: 1998-05-07
 ; PRIOR APPLICATION NUMBER: 60/084627
 ; PRIOR FILING DATE: 1998-05-07
 ; PRIOR APPLICATION NUMBER: 60/084637
 ; PRIOR FILING DATE: 1998-05-07
 ; PRIOR APPLICATION NUMBER: 60/085149
 ; PRIOR FILING DATE: 1998-05-12
 ; PRIOR APPLICATION NUMBER: 60/085323
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: 60/085338
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: 60/085339
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: 60/085579
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/085697
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/085704
 ; PRIOR FILING DATE: 1998-05-15
 ; PRIOR APPLICATION NUMBER: 60/086414
 ; PRIOR FILING DATE: 1998-05-22
 ; PRIOR APPLICATION NUMBER: 60/086430
 ; PRIOR FILING DATE: 1998-05-22
 ; PRIOR APPLICATION NUMBER: 60/087106
 ; PRIOR FILING DATE: 1998-05-28
 ; PRIOR APPLICATION NUMBER: 60/088026
 ; PRIOR FILING DATE: 1998-06-04
 ; PRIOR APPLICATION NUMBER: 60/088730
 ; PRIOR FILING DATE: 1998-06-10
 ; PRIOR APPLICATION NUMBER: 60/088741
 ; PRIOR FILING DATE: 1998-06-10
 ; PRIOR APPLICATION NUMBER: 60/088810
 ; PRIOR FILING DATE: 1998-06-10
 ; PRIOR APPLICATION NUMBER: 60/088858
 ; PRIOR FILING DATE: 1998-06-11
 ; PRIOR APPLICATION NUMBER: 60/089532
 ; PRIOR FILING DATE: 1998-06-17
 ; PRIOR APPLICATION NUMBER: 60/089599
 ; PRIOR FILING DATE: 1998-06-17
 ; PRIOR APPLICATION NUMBER: 60/089907
 ; PRIOR FILING DATE: 1998-06-18
 ; PRIOR APPLICATION NUMBER: 60/089947
 ; PRIOR FILING DATE: 1998-06-19
 ; PRIOR APPLICATION NUMBER: 60/090349
 ; PRIOR FILING DATE: 1998-06-23
 ; PRIOR APPLICATION NUMBER: 60/090429
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090445
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090538
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090863
 ; PRIOR FILING DATE: 1998-06-26
 ; PRIOR APPLICATION NUMBER: 60/091360
 ; PRIOR FILING DATE: 1998-07-01
 ; PRIOR APPLICATION NUMBER: 60/091519
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091982
 ; PRIOR FILING DATE: 1998-07-07

Query Match 89.3%; Score 13.4; DB 9; Length 3355;
 Best Local Similarity 93.3%; Pred. No. 8.6e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAATGAC 15

DB 815 AGCCATAATGAC 829
 Search completed: June 26, 2003, 04:57:26
 Job time : 66.2 secs


```

/db_xref="taxon:8090"
/clone="MF01SSB020H21"
/clone_lib="MF01SSB cDNA"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
BASE COUNT      31 a      33 c      22 g      22 t      1 others
ORIGIN

Query Match      100.0%; Score 15; DB 13; Length 109;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCCATACATGAC 15
        |||
        16 AGCCATACATGAC 30

RESULT 2
AO931789
LOCUS
DEFINITION      AO931789      396 bp      DNA      linear      GSS 21-DEC-1999
                  RPCI-23-269112.TV RPCI-23 Mus musculus genomic clone RPCI-23-269112
                  , DNA sequence.
ACCESSION
VERSION          AO931789.1 GI:6620803
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 396)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Aklnet
,B., Levins,M., McGinn,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-269112.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(piet@edjong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 269 row: I column: 12
Seq primer: 77
Class: BAC ends.
FEATURES
Source
Location/Qualifiers
1..396
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-269112"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;
ECORI: Site 2; EcorI: Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcorI and EcorI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      102 a      105 c      75 g      114 t
ORIGIN

Query Match      100.0%; Score 15; DB 17; Length 396;

```

```

Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCCATACATGAC 15
        |||
        5 AGCCATACATGAC 19

Db      5 AGCCATACATGAC 19

RESULT 3
AO664590/c
LOCUS
DEFINITION      AO664590      473 bp      DNA      linear      GSS 23-JUN-1999
                  HS.5220.B1.H03.SP6E RPCI-11 Human Male BAC Library Homo sapiens
                  genomic clone Plate=796 Col=5 Row=P, DNA sequence.
ACCESSION
VERSION          AO664590.1 GI:5172358
KEYWORDS
SOURCE
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 473)
Mahaliras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahaliras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piet@edjong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 796 row: P column: 5
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 473.
FEATURES
Source
Location/Qualifiers
1..473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=796 Col=5 Row=P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site: 1; EcorI: Site 2; EcorI:
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcorI and
EcorI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcorI sites"
BASE COUNT      134 a      89 c      95 g      135 t
ORIGIN

Query Match      100.0%; Score 15; DB 17; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCCATACATGAC 15
        |||
        354 AGCCATACATGAC 340

Db      354 AGCCATACATGAC 340

RESULT 4
AO312321/c
LOCUS
DEFINITION      AO312321      539 bp      DNA      linear      GSS 04-MAY-1999
                  RPCI11-112M12.TV RPCI-11 Homo sapiens genomic clone RPCI-11-112M12,

```


DNA sequence.
 ACCESSION A0312321
 VERSION A0312321.1 GI:4043985
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 539)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
 TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
 JOURNAL Unpublished (1998)
 COMMENT Other GSSs: RPC11-112M12.TV
 Contact: Shaying Zhao, William Niernan, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..539
 /organism="Homo sapiens"
 /db_xref="GDB:7542923"
 /db_xref="taxon:9606"
 /clone="RPC1-11-112M12"
 /clone_1lb="RPC1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site1: EcoRI; Site2: EcoRI; RPC11 Human Male BAC Library"
 BASE COUNT 163 a 97 c 106 g 173 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 17; Length 539;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AGCCATACATGAC 15
 |||||||||||||
 Db 357 AGCCATACATGAC 343
 RESULT 5
 A2620907 609 bp DNA linear GSS 13-DEC-2000
 LOCUS 1M0453M18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0453M18 R, DNA sequence.
 ACCESSION A2620907
 VERSION A2620907.1 GI:11743097
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 609)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss

University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0453 row: M column: 18
 Seq primer: CACACAGGAACACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 609.
 FEATURES
 source Location/Qualifiers
 1..609
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0453M18"
 /clone_1lb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42ny. Purified genomic DNA from Musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b/A129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
 BASE COUNT 192 a 126 c 113 g 178 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 17; Length 609;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AGCCATACATGAC 15
 |||||||||||||
 Db 144 AGCCATACATGAC 158
 RESULT 6
 BH616021/c 637 bp DNA linear GSS 28-JAN-2002
 LOCUS BMBAC305H05SP6_P5U Brugia malayi genomic Bac Library 3 Brugia
 DEFINITION malayi genomic, DNA sequence.
 ACCESSION BH616021
 VERSION BH616021.1 GI:18380709
 KEYWORDS GSS.
 SOURCE Brugia malayi.
 ORGANISM Brugia malayi
 Eukaryota; Metazoa; Nematoda; Chromadorea; Splturida; Filarioidea; Onchocercidae; Brugia.
 REFERENCE 1 (bases 1 to 637)
 AUTHORS Whitton,C., Dab,J., Ware,J., Quail,M., Hall,N., Barrell,B., Foster,J., Guillano,D., Slatko,B. and Blaxter,M.
 TITLE Genome survey sequences from the human parasitic nematode Brugia malayi
 JOURNAL Unpublished (2000)
 COMMENT Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh

Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 Sequenced from the Brugia malayi BAC library constructed by Claire Whitton and Dr Mike Quail. The sequence was generated by The Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh, Edinburgh, UK
 Seq primer: SP6 (ATTAGGTGACACTATAG)
 Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..637

/organism="Brugia malayi"

/strain="T8S"

/db_xref="taxon:6279"

/clone_lib="Brugia malayi Genomic Bac Library 3"

/sex="Mixed (male and female)"

/tissue_type="whole parasite"

/dev_stage="microfilaria (L1)"

/note="Vector: pBACE3.6; Site_1: BamH I; Brugia malayi genomic DNA was partially cleaved with Sau3A I and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitton, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."

BASE COUNT

196 a 109 c 120 g 212 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 15; DB 17; Length 637;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 AGCCATACATGAC 15

Db

489 AGCCATACATGAC 475

RESULT 7

A2572103/c

LOCUS A2572103 657 bp DNA linear GSS 15-MAY-2001

DEFINITION 303pVD05 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.

ACCESSION A2572103

VERSION A2572103.1 GI:13985046

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

1 (bases 1 to 657)

The Plasmodium vivax and P. berghoi gene sequence tag projects

Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..657

/organism="Plasmodium vivax"

/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"

/db_xref="taxon:5855"

/clone_lib="PV MBN #30"

/dev_stage="asexual blood forms"

/lab_host="Saimiri boliviensis"

/note="Vector: pBluescript SK(+); vector DNA, phagemid

excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmidipur filter, followed by passage through a column of pre-wet Whatman CF11 powder (1:2 ratio volume of blood to CF11), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 500C as described (Verneick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

BASE COUNT

147 a 146 c 148 g 213 t 3 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 15; DB 17; Length 657;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 AGCCATACATGAC 15

Db

294 AGCCATACATGAC 280

RESULT 8

BM038276/c

LOCUS BM038276 724 bp mRNA linear EST 06-NOV-2001

DEFINITION U006D06 Oryza sativa mature leaf library induced by M.grisea Oryza

sativa cDNA clone U006D06, mRNA sequence.

ACCESSION BM038276

VERSION BM038276.1 GI:16753897

KEYWORDS

SOURCE

ORGANISM

Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 724)

A Gene Expression Screen in Oryza sativa

Unpublished (2001)

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..724

/organism="Oryza sativa"

/db_xref="taxon:4530"

/clone="U006D06"

/clone_lib="Oryza sativa mature leaf library induced by

M.grisea"

/tissue_type="leaf"

/dev_stage="Mature stage"

/note="Vector: pSport2"

BASE COUNT

149 a 181 c 185 g 206 t 3 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 15; DB 13; Length 724;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAC 15
|||||
Db 53 AGCCATACATGAC 39

RESULT 9
AG123454 754 bp DNA linear GSS 04-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-133B19.R, genomic survey sequence.
DEFINITION AG123454
ACCESSION AG123454.1 GI:16652619
VERSION GSS.
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE BAC library clone:PTB-133B19.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 754)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-cho,Tsurtumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS

FEATURES
source
LIBRARY Sequencing: M13Rev
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..754
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-133B19.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 258 a 124 c 127 g 245 t
ORIGIN

Query Match 100.0%; Score 15; DB 17; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAC 15
|||||
Db 585 AGCCATACATGAC 571

RESULT 10
CNS04EDM 891 bp DNA linear GSS 21-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 104K08 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL286925.1 GI:8025396
VERSION GSS; genome survey sequence.
KEYWORDS Tetradon nigroviridis.
SOURCE Tetradon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE 1
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brotier,P., Queller,F.,
Saurin,W. and Weissbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
JOURNAL Tetradon nigroviridis DNA sequence
REFERENCE 2 (bases 1 to 891)
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billaule,A., Queller,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 891)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
location/Qualifiers
1..891
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="104K08"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG104BF04SP1-end :
PUC-ori"

BASE COUNT 164 a 252 c 247 g 211 t 17 others
ORIGIN

Query Match 100.0%; Score 15; DB 17; Length 891;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAC 15
|||||
Db 490 AGCCATACATGAC 476

RESULT 11
BE326934 288 bp mRNA linear EST 14-JUL-2000
LOCUS hr67g09.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:313600 3',
DEFINITION mRNA sequence.
ACCESSION BE326934
VERSION BE326934.1 GI:9200710
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 288)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNT, send email to:
info@image.llnl.gov
Seq primer: 400P from Gibco.
location/Qualifiers

```

source
1..288
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1333600"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT      102 a      70 c      30 g      86 t
ORIGIN
Query Match      93.3%; Score 14; DB 10; Length 288;
Best Local Similarity 100.0%; Pred. No. 3.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCATACATGAA 14
        |||||||
Db      191 AGCCATACATGAA 204

RESULT 12
LOCUS      AA953965      350 bp      mRNA      linear      EST 07-JUL-1998
DEFINITION o088906.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1573306 3'
            similar to gp:X56740 RAS-RELATED PROTEIN RAB-11 (HUMAN);. mRNA
            sequence.
ACCESSION  AA953965
VERSION     AA953965.1 GI:3116883
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 350)
AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLW at:
            www.bio.linn.gov/dbp/imag/image.html
            Insert Length: 952 Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 147.
            Location/Qualifiers
                1..350
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1573306"
                /clone_lib="NCI_CGAP_Kid5"
                /tissue_type="2 pooled tumors (clear cell type)"
                /lab_host="DH10B"
                /note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
                a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer [5'
                AACGCAAGATTGCGCGCCCAATATTTTATTTTATTTT 3'],
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I

```

```

and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo.
BASE COUNT      127 a      60 c      72 g      91 t
ORIGIN
Query Match      93.3%; Score 14; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCATACATGAA 14
        |||||||
Db      35 AGCCATACATGAA 48

RESULT 13
LOCUS      AW433346/c      381 bp      mRNA      linear      EST 03-DEC-2001
DEFINITION sh55h12.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1015-3744 5' similar to TR:P96167 P96167 PUTATIVE ALDOLASE. ;,
            mRNA sequence.
ACCESSION  AW433346
VERSION     AW433346.1 GI:6964653
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
            1 (bases 1 to 381)
REFERENCE  1
AUTHORS   Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corvelli,V., Khanna
            A., Bolla,B., Merris,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            Y., Persson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            R., Rittler,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            R., Waterston,R. and Wilson,R.
            Public Soybean EST Project
            Unpublished (1999)
JOURNAL    Contact: Shoemaker R./Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: c@resgen.com
            Insert Length: 730 Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 336.
            Location/Qualifiers
                1..381
                /organism="Glycine max"
                /db_xref="taxon:3847"
                /clone="GENOME SYSTEMS CLONE ID: Gm-c1015-3744"
                /clone_lib="Gm-c1015"
                /tissue_type="Mature flowers, field grown plants"
                /lab_host="XL10-Gold"
                /note="Vector: pBluescript II Xr; Site_1: EcoRI; Site_2:
                XhoI; This cDNA library was constructed from mRNA isolated
                from mature flowers of field grown plants. The cDNA
                library was prepared using the Stratagene pBluescript II
                XR cDNA library construction kit. Complementary DNA was
                synthesized from mRNA using a primer consisting of a poly
                (dT) sequence with a XhoI restriction site. EcoRI adaptors
                were ligated to the blunt-ended cDNA fragments followed by
                XhoI digestion. The cDNA fragments were directionally
                cloned into the EcoRI-XhoI restriction site of the
                pBluescript vector. The ligated cDNA fragments were
                transformed into XL10-Gold host cells. This library was
                constructed by Dr. Randy Shoemaker and Dr. John

```

Erpeliding."

BASE COUNT 109 a 67 c 98 g 107 t

ORIGIN

Query Match 93.3%; Score 14; DB 10; Length 381;
 Best Local Similarity 100.0%; Pred. No. 4.2e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GCCATACATGAAC 15
 |||||||
 Db 314 GCCATACATGAAC 301

RESULT 14
 BQ048783 384 bp mRNA linear EST 29-MAR-2002
 LOCUS 952024A09.y2 952 - BMS tissue from Walbot Lab (reduced rRNA) Zea
 DEFINITION
 mays cDNA, mRNA sequence.
 ACCESSION BQ048783
 VERSION BQ048783.1 GI:19822759
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 384)
 Walbot,V.
 Maize ESTs from various cDNA libraries sequenced at Stanford University

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 952024 row: A column: 09.

FEATURES
 Source
 Location/Qualifiers

1..384
 /organism="Zea mays"
 /cultivar="BMS (Black Mexican Sweet)"
 /db_xref="taxon:4577"
 /clone_lib="952 - BMS tissue from Walbot Lab (reduced rRNA)"
 /tissue_type="suspension culture"
 /dev_stage="mixed logarithmic and stationary growth phases"
 /lab_host="DH10B"
 /note="Vector: pUC19; Site_1: EcoRI; Site_2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

BASE COUNT 91 a 102 c 62 g 129 t

ORIGIN

Query Match 93.3%; Score 14; DB 14; Length 384;
 Best Local Similarity 100.0%; Pred. No. 4.2e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAA 14
 |||||||
 Db 358 AGCCATACATGAA 371

RESULT 15

BH863054/C
 LOCUS 403 bp DNA linear GSS 05-AUG-2002
 DEFINITION SALK_093049 Arabidopsis thaliana TDNA insertion lines Arabidopsis
 thaliana genomic clone SALK_093049, DNA sequence.

ACCESSION BH863054
 VERSION BH863054.1 GI:22098383
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 403)
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab,
 C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
 Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 Unpublished (2001)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

The Salk Institute Genomic Analysis Laboratory (SIGAL)
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA.

FEATURES
 source
 Location/Qualifiers

1..403
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_093049"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tdna-protocols.html>

BASE COUNT 92 a 102 c 84 g 125 t

ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 403;
 Best Local Similarity 100.0%; Pred. No. 4.3e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GCCATACATGAAC 15
 |||||||
 Db 245 GCCATACATGAAC 232

Search completed: June 26, 2003, 04:50:18
 Job time : 753.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds

(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-42

Perfect score: 15

Sequence: 1 agttactactcgagc 15

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBankl:*
1: gb_ba:*
2: gb_hc:*
3: gb_in:*
4: gb_com:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_com:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	6	AR183495
2	15	100.0	209	9	HS1GHWG41
3	15	100.0	218	9	HSJ227742
4	15	100.0	219	9	HSU086367
5	15	100.0	220	9	HSJ227729
6	15	100.0	221	9	HSVH4GL7
7	15	100.0	225	9	S75891
8	15	100.0	229	9	HSJA347134
9	15	100.0	230	9	HSVH4GL11
10	15	100.0	234	9	HSVH4GL10
11	15	100.0	234	9	HSVH4GL9
12	15	100.0	235	9	HSVH4GL1
13	15	100.0	236	9	HS1G2B1
14	15	100.0	237	9	HSRDP71RB
15	15	100.0	238	9	HSJA347131
16	15	100.0	238	9	HSJA347367
17	15	100.0	238	9	HSJA347368
18	15	100.0	240	9	HSJA306255
19	15	100.0	240	9	HSJA347122
20	15	100.0	241	9	HSJA347352
21	15	100.0	244	9	HSJA347166
22	15	100.0	248	9	S74500
23	15	100.0	249	9	AY013307
24	15	100.0	251	9	HSJA347242
25	15	100.0	253	9	HSJA298472
26	15	100.0	253	9	HSJA14778
27	15	100.0	256	9	HSJA402538
28	15	100.0	256	9	HSVIGD2
29	15	100.0	258	9	HSJA402422
30	15	100.0	261	9	HSJA402380
31	15	100.0	262	9	HSJA14770
32	15	100.0	262	9	S74642
33	15	100.0	264	9	HSSEM16V4
34	15	100.0	265	9	HSJA402546
35	15	100.0	267	9	HSVIGD4
36	15	100.0	267	9	HSVIGD6
37	15	100.0	268	9	HSJA347149
38	15	100.0	268	9	HSJA402470
39	15	100.0	270	9	HS1G18V4
40	15	100.0	270	9	HS280554
41	15	100.0	272	9	HSJA402535
42	15	100.0	273	9	HS1G31V4
43	15	100.0	276	9	AF077488
44	15	100.0	276	9	HSVCBG9
45	15	100.0	276	9	HS280734

ALIGNMENTS

RESULT 1
AR183495
LOCUS AR183495 15 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 42 from patent US 6342220.
ACCESSION AR183495
VERSION AR183495.1 GI:20227464
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 42 29-JAN-2002;
FEATURES Location/Qualifiers

```

source          1..15
                /organism="unknown"
BASE COUNT      4 a          3 c          4 g          4 t
ORIGIN

Query Match      100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY              1 AGTTACTACTGGAGC 15
                |||
Db              1 AGTTACTACTGGAGC 15

RESULT 2
LOCUS          HSIGVHG41
DEFINITION     H.sapiens germline immunoglobulin heavy chain, variable region,
                (G411).
ACCESSION      X92296
VERSION         X92296.1 GI:1045179
KEYWORDS       germ line; immunoglobulin.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE          1 (bases 1 to 209)
JOURNAL        Ikematsu,H., Kasaiian,M.T., Schettino,E.W. and Casali,P.
                Structural analysis of the VH-D-JH segments of human polyclonal
                196 mAb. Evidence for somatic selection
                J. Immunol. 151 (7), 3604-3616 (1993)
MEDLINE        93389188
PUBMED         8376796
FEATURES       Location/Qualifiers
                1..209
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT     49 a          61 c          56 g          43 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 209;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY              1 AGTTACTACTGGAGC 15
                |||
Db              27 AGTTACTACTGGAGC 41

RESULT 3
LOCUS          HSJ227742
DEFINITION     Homo sapiens DNA for rearranged immunoglobulin heavy chain gene,
                5-2VH4.
ACCESSION      AJ227742
VERSION         AJ227742.1 GI:2951632
KEYWORDS       heavy chain; immunoglobulin; variable region.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE          1 (bases 1 to 218)
JOURNAL        Delabie,J.
                Direct Submission
                Submitted (03-MAR-1998) Delabie J., Pathology, University of
                Leuven, Minderbroedersstraat 12, B-3000, BELGIUM
REFERENCE       2 (bases 1 to 218)
AUTHORS        Delabie,J. and Tiersens,A.
TITLE          Evidence for clonal expansion and somatic hypermutations of the
                marginal zone B cells in the lymph node and spleen
                Unpublished
                Location/Qualifiers
                1..218

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/rearranged
<1..>218
/codon_start=2
/product="Immunoglobulin heavy chain variable region"
/protein_id="CAI12720.1"
/db_xref="GI:2951633"
/translation="GGSVSSYYWSWIRQPPGKLEWIGYIFYTGSTNYPNPSLSRVTI
SIDMSTQDFSLKISSVTADTAIVYCAR"
<1..>218
/product="Immunoglobulin heavy chain variable region"
BASE COUNT     48 a          61 c          60 g          49 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY              1 AGTTACTACTGGAGC 15
                |||
Db              17 AGTTACTACTGGAGC 31

RESULT 4
LOCUS          HSU86367
DEFINITION     Human rearranged Igh variable region gene, partial cds.
ACCESSION      U86367
VERSION         U86367.1 GI:1841799
KEYWORDS
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE          1 (bases 1 to 219)
JOURNAL        Tiersens,A. and Delabie,J.
                Marginal zone lymphoma 8-VH4
                Unpublished
                2 (bases 1 to 219)
REFERENCE       Tiersens,A. and Delabie,J.
                Direct Submission
                Submitted (22-JAN-1997) Pathology, University of Leuven,
                Minderbroedersstraat 12, Leuven B-3000, Belgium
                Location/Qualifiers
                1..219
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /rearranged
                /note="marginal zone lymphoma 8-VH4"
                <1..>219
                /codon_start=1
                /product="Igh variable region"
                /protein_id="AAB47529.1"
                /db_xref="GI:1841800"
                /translation="SGGSISYYWSWIRQPPGKLEWIGYIFYTGSTNYPNPSLSRVT
                ISVPTSKNQSFKINSVIAADTAIVYCAR"
BASE COUNT     53 a          54 c          56 g          56 t
ORIGIN

Query Match      100.0%; Score 15; DB 9; Length 219;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY              1 AGTTACTACTGGAGC 15
                |||
Db              19 AGTTACTACTGGAGC 33

RESULT 5
LOCUS          HSJ227729
DEFINITION     Homo sapiens DNA for rearranged immunoglobulin heavy chain gene,

```



```

2-11VH4.
AJ227729
VERSION AJ227729.1 GI:2951609
KEYWORDS heavy chain; immunoglobulin; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Delabie,J.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-1998) Delabie J., Pathology, University of
Leuven, Minderbroederstraat 12, B-3000, BELGIUM
REFERENCE
AUTHORS Delabie,J. and Trierens,A.
TITLE Evidence for clonal expansion and somatic hypermutations of the
marginal zone B cells in the lymph node and spleen
JOURNAL Unpublished
FEATURES
source Location/Qualifiers
1..220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/rearranged
<1..>220
/codon_start=3
/product="immunoglobulin heavy chain variable region"
/db_xref="GI:2951610"
/translation="GGSTISYWSWIRPPCKGLEWIGIYISGSGTNYNPGLSKQVTI
SVDTSKNPFSLKSSVTADPAAVYCAR"
<1..>220
/product="immunoglobulin heavy chain variable region"

V_region
BASE COUNT 54 a 60 c 60 g 46 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
|||||
18 AGTTACTACTGAGC 32

RESULT 6
LOCUS HSYHAGL7 221 bp DNA linear PRI 27-JUN-1996
DEFINITION H.sapiens germline immunoglobulin heavy chain VH gene (clone
VH4-GL7).
ACCESSION 275359
VERSION 275359.1 GI:1418960
KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin
superfamily; immunoglobulin variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Voswinkel,J., Triemper,L., Carbon,G., Hopf,T., Pfeundschuh,M. and
Gause,A.
TITLE Evidence for a selected humoral immune response encoded by VH4
family genes in the synovial membrane of a patient with rheumatoid
arthritis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 221)
AUTHORS Voswinkel,J.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland
Medical School, D-66421 Homburg/Saar, GERMANY
FEATURES
source Location/Qualifiers
1..221
/organism="Homo sapiens"
/isolate="patient 8 with rheumatoid arthritis"

```

```

/db_xref="taxon:9606"
/chromosome="14"
/clone="VH4-GL7"
/cell_type="B_lymphocyte"
/tissue_type="synovial membrane"
/clone_id="VH4-GL"
/germline

V_segment
BASE COUNT 51 a 60 c 59 g 47 t 4 others
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 221;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
|||||
19 AGTTACTACTGAGC 33

RESULT 7
LOCUS S75891 225 bp mRNA linear PRI 27-JUL-1995
DEFINITION IgH-immunoglobulin heavy chain variable region [clone M1] [human,
multiple myeloma patient, mRNA partial, 225 nt].
ACCESSION S75891
VERSION S75891.1 GI:913791
KEYWORDS
SOURCE Homo sapiens multiple myeloma patient.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Baker,B.W., Deane,M., Gilleece,M.H., Johnston,D., Scarffe,J.H. and
Norton,J.D.
TITLE Distinctive features of immunoglobulin heavy chain variable region
gene rearrangement in multiple myeloma
JOURNAL Leuk. Lymphoma 14 (3-4), 291-301 (1994)
MEDLINE 95038365
PUBMED 7950918
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gdbseq 161439] from the original journal article.
FEATURES
source Location/Qualifiers
1..225
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..225
/partial
/gene="IgH"
/note="immunoglobulin heavy chain variable region"
1..225
/gene="IgH"
/note="immunoglobulin heavy chain variable region"
/codon_start=1
/protein_id="AADI4200.1"
/db_xref="GI:4261900"
/translation="AVSGYSSYWSWIRPPCKGLEWIGIYISGSGTNYNPGLSKSR
ATISVDTSKNPFSLKSSVTADPAAVYCAR"

BASE COUNT 48 a 67 c 56 g 54 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
|||||
25 AGTTACTACTGAGC 39

RESULT 8
LOCUS HSA347134

```

LOCUS HSA347134 229 bp DNA linear PRI 26-MAR-2002
 DEFINITION Homo sapiens partial IGWH4-61 gene for immunoglobulin heavy chain variable region, isolate case2-cell1141.
 ACCESSION AJ347134
 VERSION AJ347134.1 GI:19773037
 KEYWORDS IGHV gene; immunoglobulin heavy chain; variable region.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
 AUTHORS Brauningner, A., Spieker, T., Willenbrock, K., Gaulard, P., Wacker, H.H., Rajewsky, K., Hansmann, M.L. and Kuppers, R. 'forbidden' Survival and clonal expansion of mutating (Immunoglobulin receptor-deficient) Epstein-Barr virus-infected B cells in angioimmunoblastic T cell lymphoma
 JOURNAL J. Exp. Med. 194 (7), 927-940 (2001)
 MEDLINE 21464858
 PUBMED 11581315
 REFERENCE 2 (bases 1 to 229)
 AUTHORS Brauningner, A.
 TITLE Direct Submission
 JOURNAL Submitted (20-AUG-2001) Brauningner A., Pathology, University of Frankfurt, Theodor-Stern-Kai 7, 60590, GERMANY

FEATURES
 source
 1..229
 /organism="Homo sapiens"
 /isolate="case2-cell1141"
 /db_xref="taxon:9606"
 /rearranged
 1..229
 /gene="IGWH4-61"
 <1..>229
 /gene="IGWH4-61"
 /codon_start=1
 /product="immunoglobulin heavy chain variable region"
 /protein_id="C687534.1"
 /db_xref="GI:19773038"
 /translation="GSYYWSWIRQPAKGLWIGRLTYSGNTNYPNLSKRVIMSYDVT SKNPSLKSISVTADPAVYVCARDLGLIDSM"
 <1..>229
 /gene="IGWH4-61"
 /product="immunoglobulin heavy chain variable region"

BASE COUNT 51 a 66 c 65 g 47 t

ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 229;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTACTACTGAGC 15
 |||||||
 4 AGTACTACTGAGC 18

DB

RESULT 9
 LOCUS HSVH4GL11 230 bp DNA linear PRI 27-JUN-1996
 DEFINITION H.sapiens germline immunoglobulin heavy chain VH gene (clone VH4-GL11).
 ACCESSION 275348
 VERSION 275348.1 GI:1418949
 KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; immunoglobulin variable region.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 230)
 AUTHORS Voswinkel, J., Truemper, L., Carbon, G., Hopf, T., Pfrendschuh, M. and Gause, A.
 TITLE Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with rheumatoid

JOURNAL arthritits
 REFERENCE 2 (bases 1 to 230)
 AUTHORS Voswinkel, J.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland Medical School, D-66421 Homburg/Saar, GERMANY

FEATURES
 source
 1..230
 /organism="Homo sapiens"
 /isolate="patient 8 with rheumatoid arthritis"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="VH4-GL11"
 /cell_type="B lymphocyte"
 /tissue_type="synovial membrane"
 /clone_lib="VH4-GL"
 /germline

BASE COUNT 54 a 64 c 64 g 48 t

ORIGIN

V segment 1..230

Query Match 100.0%; Score 15; DB 9; Length 230;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTACTACTGAGC 15
 |||||||
 25 AGTACTACTGAGC 39

DB

RESULT 10
 LOCUS HSVH4GL10 234 bp DNA linear PRI 27-JUN-1996
 DEFINITION H.sapiens germline immunoglobulin heavy chain VH gene (clone VH4-GL10).
 ACCESSION 275347
 VERSION 275347.1 GI:1418948
 KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; immunoglobulin variable region.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 234)
 AUTHORS Voswinkel, J., Truemper, L., Carbon, G., Hopf, T., Pfrendschuh, M. and Gause, A.
 TITLE Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with rheumatoid arthritis

JOURNAL unpublished
 REFERENCE 2 (bases 1 to 234)
 AUTHORS Voswinkel, J.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland Medical School, D-66421 Homburg/Saar, GERMANY

FEATURES
 source
 1..234
 /organism="Homo sapiens"
 /isolate="patient 8 with rheumatoid arthritis"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="VH4-GL10"
 /cell_type="B lymphocyte"
 /tissue_type="synovial membrane"
 /clone_lib="VH4-GL"
 /germline

BASE COUNT 56 a 65 c 65 g 48 t

ORIGIN

V segment 1..234

Query Match 100.0%; Score 15; DB 9; Length 234;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
 |||||||
 Db 25 AGTTACTACTGAGC 39

RESULT 11
 HSVH4GL9 234 bp DNA linear PRI 27-JUN-1996
 LOCUS H.sapiens germline immunoglobulin heavy chain VH gene (clone
 DEFINITION VH4-GL9).
 ACCESSION 275361
 VERSION 275361.1 GI:1418962
 KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin
 superfamily; immunoglobulin variable region.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 234)
 Voswinkel,J., Truemper,L., Carbon,G., Hopf,T., Pfrendschuh,M. and
 Cause,A.
 Evidence for a selected humoral immune response encoded by VH4
 family genes in the synovial membrane of a patient with rheumatoid
 arthritis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 234)
 AUTHORS Voswinkel,J.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland
 Medical School, D-66421 Homburg/Saar, GERMANY
 FEATURES
 source Location/Qualifiers
 1..234
 /organism="Homo sapiens"
 /isolate="patient 8 with rheumatoid arthritis"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="VH4-GL9"
 /cell_type="B lymphocyte"
 /tissue_type="synovial membrane"
 /clone_id="VH4-GL"
 /germline
 V_segment 1..234
 BASE COUNT 56 a 65 c 65 g 48 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 9; Length 234;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
 |||||||
 Db 25 AGTTACTACTGAGC 39

RESULT 12
 HSHV4GL1 235 bp DNA linear PRI 27-JUN-1996
 LOCUS H.sapiens immunoglobulin heavy chain VH gene (clone VH4-GL1).
 DEFINITION 275346
 ACCESSION 275346.1 GI:1418947
 VERSION 275346.1
 KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin
 superfamily; immunoglobulin variable region.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 235)
 Voswinkel,J., Truemper,L., Carbon,G., Hopf,T., Pfrendschuh,M. and
 Cause,A.
 Evidence for a selected humoral immune response encoded by VH4
 family genes in the synovial membrane of a patient with rheumatoid

arthritis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 235)
 AUTHORS Voswinkel,J.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland
 Medical School, D-66421 Homburg/Saar, GERMANY
 FEATURES
 source Location/Qualifiers
 1..235
 /organism="Homo sapiens"
 /isolate="patient 8 with rheumatoid arthritis"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="VH4-GL1"
 /cell_type="B lymphocyte"
 /tissue_type="synovial membrane"
 /clone_id="VH4-GL"
 /germline
 V_segment 1..235
 BASE COUNT 55 a 64 c 67 g 49 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 9; Length 235;
 Best Local Similarity 100.0%; Pred. No. 6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
 |||||||
 Db 26 AGTTACTACTGAGC 40

RESULT 13
 HSHGH2B1 236 bp DNA linear PRI 07-JUL-1998
 LOCUS H.sapiens immunoglobulin rearranged heavy chain gene DPe6 (clone
 DEFINITION 2B1).
 ACCESSION 271777
 VERSION 271777.1 GI:1279486
 KEYWORDS diversity region; immunoglobulin; immunoglobulin heavy chain;
 joining region; variable region.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 236)
 Dunn-Walters,D.K.
 Direct Submission
 JOURNAL Submitted (19-APR-1996) Dunn-Walters D.K., University College
 Medical School, Department of Histopathology, Rockefeller Building,
 University Street, London, WC1E 6JJ, UK
 2 (bases 1 to 236)
 Dunn-Walters,D.K., Boursier,L., Spencer,J. and Isaacson,P.G.
 Analysis of immunoglobulin genes in splenic marginal zone lymphoma
 suggests ongoing mutation
 Hum. Pathol. 29 (6), 585-593 (1998)
 JOURNAL 98297601
 MEDLINE 9635678
 PUBMED
 FEATURES
 source Location/Qualifiers
 1..236
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="2B1"
 /tissue_type="spleen (splenic marginal zone lymphoma)"
 /rearranged
 BASE COUNT 51 a 69 c 63 g 53 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 9; Length 236;
 Best Local Similarity 100.0%; Pred. No. 6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15

```

Db          43 AGTTACTACTGGAGC 57
|||||
RESULT 14
HSDP71RB      237 bp DNA linear PRI 03-OCT-1996
LOCUS         H.sapiens B9-g1DP71RB germ-line gene for immunoglobulin heavy chain
DEFINITION    Variable region.
ACCESSION     X87091.1 GI:1592730
VERSION       X87091.1
KEYWORDS      germ-line gene; immunoglobulin; immunoglobulin heavy chain; variable
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE     1 (bases 1 to 237)
AUTHORS       Brezinschek, H.P., Brezinschek, R.I. and Lipsky, P.E.
TITLE         Analysis of the heavy chain repertoire of human peripheral B cells
              using single-cell polymerase chain reaction
JOURNAL       J. Immunol. 155 (1), 190-202 (1995)
MEDLINE       95325588
PUBMED        7602095
REFERENCE     2 (bases 1 to 237)
AUTHORS       Brezinschek, H.P.
TITLE         Direct Submission
JOURNAL       Submitted (05-MAY-1995) Brezinschek H.P., Department of Internal
              Medicine, Harold C. Simmons Arthritis Research Center, University
              of Texas Southwestern Medical Center, 5323 Harry Hines Blvd.,
              Dallas, TX 75235-8884, USA
              On Oct 4, 1996 this sequence version replaced gi:1052697.
FEATURES
  source       1..237
               /organism="Homo sapiens"
               /isolate="donor RB"
               /db_xref="taxon:9606"
               /chromosome="14"
               /cell_type="granulocyte"
               /tissue_type="blood"
               1..237
               /gene="B9-g1DP71RB"
               /gene="B9-g1DP71RB"
               /note="VH4 family germ-line gene, polymorphic form of
               DP-71"
  BASE COUNT   54 a 70 c 61 g 50 t 2 others
  ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 237;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
   |||||||
Db  37 AGTTACTACTGGAGC 51

RESULT 15
HSA347131      238 bp DNA linear PRI 26-MAR-2002
LOCUS         Homo sapiens partial IGWH4-61 gene for immunoglobulin heavy chain
DEFINITION    Variable region, isolate case2-cell1150.
ACCESSION     AJ347131
VERSION       AJ347131.1 GI:19773031
KEYWORDS      IGWH gene; immunoglobulin heavy chain; variable region.
SOURCE        human.
ORGANISM      Homo sapiens
REFERENCE     1
AUTHORS       Brauning, A., Spicker, T., Willenbrock, K., Gaulard, P., Wacker, H.H.,
              Rejewska, K., Hansmann, M.L. and Kuppers, R.

```

```

TITLE          Survival and clonal expansion of mutating 'forbidden'
              (immunoglobulin receptor-deficient) Epstein-Barr virus-infected B
              cells in angioimmunoblastic T cell lymphoma
JOURNAL        J. Exp. Med. 194 (7), 927-940 (2001)
MEDLINE        21464858
PUBMED         11581315
REFERENCE     2 (bases 1 to 238)
AUTHORS       Brauning, A.
TITLE         Direct Submission
JOURNAL       Submitted (20-AUG-2001) Brauning, A., Pathology, University of
              Frankfurt, Theodor-Stern-Kai 7, 60590, GERMANY
FEATURES
  source       1..238
               /organism="Homo sapiens"
               /isolate="case2-cell1150"
               /db_xref="taxon:9606"
               /rearranged
               1..238
               /gene="IGWH4-61"
               <1..>238
               /gene="IGWH4-61"
               /codon_start=1
               /product="immunoglobulin heavy chain variable region"
               /protein_id="CAC87531.1"
               /db_xref="GI:19773032"
               /translation="SYMSWTRPAGKGLWIGRLTSGSTNYKSKSRVTMSVDTS
               KQPSLKSSTVADTAVYTCARGTGLRRMAFDIW"
               <1..>238
               /gene="IGWH4-61"
               /product="immunoglobulin heavy chain variable region"
  BASE COUNT   54 a 61 c 74 g 49 t
  ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
   |||||||
Db  1 AGTTACTACTGGAGC 15

```

Search completed: June 26, 2003, 03:47:37
Job time : 425.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agtactactgagac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	15	20	AA32408
2	15	100.0	110	24	ABK71391
3	15	100.0	171	22	ABA70643
4	15	100.0	171	22	ABA37202
5	15	100.0	171	22	AAK18889
6	15	100.0	171	22	AAK48833
7	15	100.0	171	22	AAI25038
8	15	100.0	171	22	AAI50810
9	15	100.0	171	24	ABSI9072

10	15	100.0	281	22	ABA70163	Human foetal liver
11	15	100.0	281	22	ABA36928	Probe #15394 for g
12	15	100.0	281	22	AAK18383	Human brain expres
13	15	100.0	281	22	AAK44283	Human bone marrow
14	15	100.0	281	22	AAI24810	Probe #14743 for g
15	15	100.0	281	22	AAI50286	Probe #18972 used
16	15	100.0	281	24	ABSI8519	Human genome-deriv
17	15	100.0	319	22	AAK20407	Human brain expres
18	15	100.0	342	24	ABK71388	Thrombopoietin (TP
19	15	100.0	348	14	AAQ42697	Vh 71-4. Homo sap
20	15	100.0	348	14	AAQ42699	VH411. Homo sapie
21	15	100.0	348	14	AAQ42700	VH415. Homo sapie
22	15	100.0	348	14	AAQ42701	VH416. Homo sapie
23	15	100.0	357	18	AAI72126	CEA-specific antib
24	15	100.0	363	22	AAH42398	Nucleotide sequenc
25	15	100.0	363	24	ABH4218	ebvHlgM MS119D10 h
26	15	100.0	369	22	AAI68756	Human autoantibody
27	15	100.0	378	22	AAH42392	Nucleotide sequenc
28	15	100.0	397	20	AAZ24417	Human bladder tumo
29	15	100.0	408	18	AAI79919	Immunoglobulin rB6
30	15	100.0	417	21	AAI13938	Human PTHrP monocl
31	15	100.0	426	24	ABK71396	DNA encoding throm
32	15	100.0	456	21	AAZ42290	Human 5' EST isola
33	15	100.0	462	22	ABA57539	Human foetal liver
34	15	100.0	462	22	ABA27008	Probe #5474 for ge
35	15	100.0	462	22	AAK05586	Human brain expres
36	15	100.0	462	22	AAK31189	Human bone marrow
37	15	100.0	462	22	AAI15585	Probe #5518 for ge
38	15	100.0	462	22	AAI37094	Probe #5780 used t
39	15	100.0	462	24	ABSI05941	Human genome-deriv
40	15	100.0	470	22	ABA58037	Human foetal liver
41	15	100.0	470	22	ABA27294	Probe #5760 for ge
42	15	100.0	470	22	AAK06111	Human brain expres
43	15	100.0	470	22	AAK31762	Human bone marrow
44	15	100.0	470	22	AAI15819	Probe #5752 for ge
45	15	100.0	470	22	AAI37638	Probe #6324 used t

ALIGNMENTS

RESULT 1	
AA332408	
ID	AA332408 standard; DNA: 15 BP.
XX	
AC	AA332408;
XX	
DT	17-JUN-1999 (first entry)
XX	
DE	Ab6 variable heavy (VH) chain CDR1 encoding DNA.
XX	
XX	
KW	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;
KW	myelodysplasia; myelotoxic chemotherapy; leukemia; tumour; MDSK; CDR;
KW	neuronmuscular; muscular dystrophy; complementarity determining region;
KW	variable heavy chain; variable light chain; VH; VL; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9910494-A2.
XX	
PD	04-MAR-1999.
XX	
EF	21-AUG-1998; 98WO-US17364.
XX	
PR	25-AUG-1997; 97US-0918148.
XX	
PA	(GETH) GENENTECH INC.
XX	
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;
XX	
DR	WPI; 1999-204666/17.

```

DR      P-PSNB, AAT06707.
PT      New thrombopoietin receptor agonist antibodies - useful for
PR      treating immunological or hematological disorders
XX
PS      Claim 10; Page 81; 86pp; English.
XX
CC      The invention relates to an agonist antibody (Ab) which binds to a
CC      thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
CC      be used in the same way and for the same indications as thrombopoietin
CC      (TPO). They can stimulate proliferation, differentiation or growth of
CC      megakaryocytes. They may also be able to stimulate megakaryocytes to
CC      increase platelet production. They can be used for treating
CC      immunological or hematopoietic disorders, especially thrombocytopenia.
CC      Following chemotherapy or bone marrow transplant (e.g. aplastic anemia
CC      treated with the antibody compounds as well as disorders such as
CC      disseminated intravascular coagulation (DIC), immune thrombocytopenia
CC      (HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
CC      congenital thrombocytopenia, thrombotic thrombocytopenia and
CC      myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
CC      treatment of solid tumours or leukaemia, myeloablative chemotherapy for
CC      autologous or allogeneic bone marrow transplant, myelodysplasia,
CC      idiopathic aplastic anemia, congenital thrombocytopenia, and immune
CC      thrombocytopenia. The antibodies which bind to the MusK receptor can be
CC      used for improving neuromuscular function in a patient, e.g. in muscular
CC      dystrophy. The products can also be used for detection and diagnosis. The
CC      antibodies have a longer half-life than the natural ligand for the TPO-R.
CC      Sequences AAx12387-X32413 represent DNA fragments encoding the CDR1,
CC      CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
CC      (VL) chains of antibodies Ab1 to Ab6.
XX
SQ      Sequence 15 BP; 4 A; 3 C; 4 G; 4 T; 0 other;
Query Match          100.0%; Score 15; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 AGTTACTACTGAGC 15
        |||
DB       1 AGTTACTACTGAGC 15
        |||
RESULT 2
ABK71391/c
ID      ABK71391 standard; DNA; 110 BP.
XX
AC      ABK71391;
XX
DT      30-JUL-2002 (first entry)
XX
DE      Thrombopoietin (TPO) agonist antibody associated polynucleotide #12.
XX
KW      Modified antibody; thrombopoietin; TPO; agonist;
XX      TPO receptor; platelet reduction-associated blood disease;
XX      thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
XX      ds.
XX
OS      Synthetic.
XX
PN      WO200233072-A1.
XX
PD      25-APR-2002.
XX
PF      22-OCT-2001; 2001WO-JP09259.
XX
PR      20-OCT-2000; 2000JP-0321821.
XX      17-APR-2001; 2001WO-JP03288.
XX      12-SEP-2001; 2001JP-0277314.
XX
PA      (CHUS ) CHUGAI SEIRAKU KK.
XX
I      Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

```

XX WPI; 2002-383513/41.
DR
XX Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
XX leukemia -
XX
PS Example 8; Page 195; 213pp; Japanese.

CC The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives
CC and/or remedies for platelet reduction-associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence represents thrombopoietin (TPO) agonist antibody associated
CC polynucleotide used in the creation of the modified antibody described in
CC the invention.
XX

SQ Sequence 110 BP; 26 A; 34 C; 31 G; 19 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 110;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 1 AGTTACTACTGGAGC 15;
|||||
Db 58 AGTTACTACTGGAGC 44

RESULT 3
ABA70643/C
ID ABA70643 standard; DNA; 171 BP.
XX
AC ABA70643;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #18948.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn Sq, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 18948; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for

measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human foetal liver. The present sequence is a single exon nucleic acid probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15
|||||
DB 82 AGTTACTACTGGAGC 68

RESULT 4
ID ABA37202 standard; DNA; 171 BP.
AC ABA37202;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #15668 for gene expression analysis in human heart cell sample.
XX
KW Human: gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human hearts -
XX
PS Claim 4; SEQ ID No 15668; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15
|||||
DB 82 AGTTACTACTGGAGC 68

RESULT 5
ID AAK18889/c
XX
AC AAK18889;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 18880.
XX
KW Human: brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human brains -
XX
PS Example 4; SEQ ID NO: 18880; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15
|||||
DB 82 AGTTACTACTGGAGC 68

RESULT 6

```
AAK44833/C
ID AAK44833 standard; DNA: 171 BP.
XX
AC AAK44833;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 19390.
XX
KM Human: bone marrow expressed exon; gene expression analysis; probe;
microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WI: 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 19390; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;
Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
Db 82 AGTTACTACTGGAGC 68
AAI25038 standard; DNA: 171 BP.
XX
AC AAI25038;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #14971 for gene expression analysis in human cervical cell sample.
XX
KM Probe; human; microarray; gene expression; cervical epithelial cell;
cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
```

```
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WI: 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 14971; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published\_pcl\_sequences.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;
Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
Db 82 AGTTACTACTGGAGC 68
AAI50810 standard; DNA: 171 BP.
XX
AC AAI50810;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #19496 used to measure gene expression in human placenta sample.
XX
KM Probe; microarray; human; placenta; antenatal diagnosis;
genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
```


PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI: 2001-488897/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 XX Claim 25; SEQ ID NO 19496; 654pp; English.
 PS
 CC The present invention relates to single exon nucleic acid probes (SRNP).
 CC The present sequence is one such probe. The probes are useful for
 CC predicting a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 XX Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;
 SQ
 Query Match 100.0%; Score 15; DB 22; Length 171;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AGTTACTACTGGAGC 15
 DB 82 AGTTACTACTGGAGC 68
 DB
 RESULT 9
 ABS19072/c
 ID ABS19072 standard; DNA; 171 BP.
 AC
 XX ABS19072;
 AC
 XX 19-AUG-2002 (first entry)
 DT
 XX Human genome-derived single exon probe ORF from lung SEQ ID NO 19063.
 DE
 XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Heremansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 KW
 OS Homo sapiens.
 OS
 XX
 XX WO200186003-A2.
 PN
 XX
 XX 15-NOV-2001.
 PD
 XX
 XX 30-JAN-2001; 2001WO-US00665.
 PF
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-063236P.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 PA
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI: 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to

PT measure gene expression in human lung samples -
 PS
 XX Claim 4; SEQ ID NO 19063; 634pp; English.
 PS
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation or detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Heremansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;
 Query Match 100.0%; Score 15; DB 24; Length 171;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AGTTACTACTGGAGC 15
 DB 82 AGTTACTACTGGAGC 68
 DB
 RESULT 10
 ABA70163
 ID ABA70163 standard; DNA; 281 BP.
 AC
 XX ABA70163;
 AC
 XX 01-FEB-2002 (first entry)
 DT
 XX Human foetal liver single exon nucleic acid probe #18468.
 DE
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200157277-A2.
 PN
 XX 09-AUG-2001.
 PD

```

XX 30-JAN-2001; 2001WO-US000669.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX
PS Claim 4; SEQ ID NO 18468; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match          100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
   |||||||
Db 64 AGTTACTACTGAGC 78

RESULT 11
ABA36928
ID ABA36928 standard; DNA; 281 BP.
XX
AC ABA36928;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #15394 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PT

```

```

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 4; SEQ ID NO 15394; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match          100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
   |||||||
Db 64 AGTTACTACTGAGC 78

RESULT 12
AAK18383
ID AAK18383 standard; DNA; 281 BP.
XX
AC AAK18383;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 18374.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.

```

XX Example 4; SEQ ID NO: 18374; 650bp + Sequence Listing; English.
PS
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;
XX
Query Match 100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGAGC 15
Db 64 AGTTACTACTGAGC 78
DB
RESULT 13
AAK44283
ID AAK44283 standard; DNA: 281 BP.
AC AAK44283;
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 18840.
DE
XX
XX Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200157276-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488900/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PT
XX
XX Example 4; SEQ ID NO: 18840; 658bp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;
XX
Query Match 100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGAGC 15
Db 64 AGTTACTACTGAGC 78
DB
RESULT 14
AA124810
ID AA124810 standard; DNA: 281 BP.
AC AA124810;
XX
XX 12-OCT-2001 (first entry)
DT
XX
XX Probe #14743 for gene expression analysis in human cervical cell sample.
DE
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488901/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PT
XX
XX Claim 25; SEQ ID No 14743; 487bp; English.
PS
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;
XX
Query Match 100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGAGC 15
Db 64 AGTTACTACTGAGC 78
DB
RESULT 15
AA150286
ID AA150286 standard; DNA: 281 BP.
XX

AC AAI50286;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #18972 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 30-JAN-2001; 2001WO-US00663.
 XX
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 18972; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 281;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
 |||||
 Db 64 AGTTACTACTGGAGC 78

Search completed: June 26, 2003, 03:11:51
 Job time : 88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(Without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agtactactgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A-COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B-COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A-COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B-COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS-COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	4	US-08-918-148-42	Sequence 42, Appl
2	15	100.0	357	2	US-08-652-816A-20	Sequence 20, Appl
3	15	100.0	622	3	US-08-545-809A-59	Sequence 59, Appl
4	15	100.0	650	3	US-08-545-809A-4	Sequence 4, Appl
5	15	100.0	1212	3	US-08-545-809A-61	Sequence 61, Appl
6	14	93.3	15	2	US-08-477-553A-3	Sequence 3, Appl
7	14	93.3	15	2	US-08-477-553A-6	Sequence 6, Appl
8	14	93.3	43	4	US-09-042-353-397	Sequence 397, Appl
9	14	93.3	43	4	US-09-042-353-400	Sequence 400, Appl
10	14	93.3	43	4	US-08-758-417A-247	Sequence 247, Appl
11	14	93.3	43	4	US-08-758-417A-250	Sequence 250, Appl
12	14	93.3	243	4	US-09-042-353-148	Sequence 148, Appl
13	14	93.3	243	4	US-08-758-417A-412	Sequence 412, Appl
14	14	93.3	282	4	US-09-042-353-149	Sequence 149, Appl
15	14	93.3	282	4	US-08-758-417A-413	Sequence 413, Appl
16	14	93.3	285	4	US-09-042-353-150	Sequence 150, Appl
17	14	93.3	285	4	US-08-758-417A-414	Sequence 414, Appl
18	14	93.3	297	4	US-09-042-353-151	Sequence 151, Appl
19	14	93.3	297	4	US-08-758-417A-415	Sequence 415, Appl
20	14	93.3	321	2	US-08-477-553A-47	Sequence 47, Appl
21	14	93.3	363	2	US-08-477-553A-50	Sequence 50, Appl
22	14	93.3	369	3	US-08-793-450-3	Sequence 3, Appl
23	14	93.3	399	3	US-08-724-752-10	Sequence 10, Appl
24	14	93.3	403	4	US-09-042-353-357	Sequence 357, Appl
25	14	93.3	403	4	US-08-758-417A-205	Sequence 205, Appl
26	14	93.3	404	4	US-09-042-353-355	Sequence 355, Appl
27	14	93.3	404	4	US-08-758-417A-203	Sequence 203, Appl

28	14	93.3	524	4	US-09-042-353-419	Sequence 419, Appl
29	14	93.3	524	4	US-08-758-417A-219	Sequence 219, Appl
30	14	93.3	631	3	US-08-545-809A-31	Sequence 31, Appl
31	14	93.3	687	3	US-08-545-809A-34	Sequence 34, Appl
32	14	93.3	1418	4	US-08-793-450-7	Sequence 7, Appl
33	14	93.3	1567	3	US-09-049-672A-17	Sequence 17, Appl
34	14	93.3	2624	4	US-09-370-838-19	Sequence 19, Appl
35	14	93.3	4926	4	US-09-042-353-418	Sequence 418, Appl
36	14	93.3	4926	4	US-08-758-417A-268	Sequence 268, Appl
37	13.4	89.3	15	2	US-08-477-553A-2	Sequence 2, Appl
38	13.4	89.3	30	1	US-08-360-125-37	Sequence 37, Appl
39	13.4	89.3	30	2	US-08-450-578-37	Sequence 37, Appl
40	13.4	89.3	30	2	US-09-017-628-37	Sequence 37, Appl
41	13.4	89.3	30	2	US-09-014-880-37	Sequence 37, Appl
42	13.4	89.3	30	4	US-08-450-363-37	Sequence 37, Appl
43	13.4	89.3	354	2	US-08-652-816A-23	Sequence 23, Appl
44	13.4	89.3	366	1	US-08-360-125-9	Sequence 9, Appl
45	13.4	89.3	366	2	US-08-450-578-9	Sequence 9, Appl

ALIGNMENTS

```
RESULT 1
US-08-918-148-42
: Sequence 42, Application US/08918148A
: Patent No. 6342220
: GENERAL INFORMATION:
: APPLICANT: Adams, Camellia
: APPLICANT: W.
: APPLICANT: Carter, Paul J.
: APPLICANT: Fendly, Brian M.
: APPLICANT: Gurney, Austin L.
: TITLE OF INVENTION: Agonist Antibodies
: FILE REFERENCE: P0979
: CURRENT APPLICATION NUMBER: US/08/918,148A
: CURRENT FILING DATE: 1997-08-25
: NUMBER OF SEQ ID NOS: 79
: SEQ ID NO 42
: LENGTH: 15
: TYPE: DNA
: ORGANISM: artificial
: FEATURE:
: NAME/KEY: 12E10scfv VH CDR1
: LOCATION: 1-15
: OTHER INFORMATION:
US-08-918-148-42
Query Match          100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      1 AGTACTACTGAGC 15
      |||
Db      1 AGTACTACTGAGC 15
      |||

RESULT 2
US-08-652-816A-20
: Sequence 20, Application US/08652816A
: Patent No. 5872215
: GENERAL INFORMATION:
: APPLICANT: Osbourne, JK
: APPLICANT: Allen, DJ
: APPLICANT: McCafferty, JG
: TITLE OF INVENTION: Specific binding members, materials and
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
```

```

? COUNTRY: United States of America
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/652,816A
? FILING DATE: 23-MAY-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9125579.4
? FILING DATE: 02-DEC-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9125579.8
? FILING DATE: 02-DEC-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9206318.9
? FILING DATE: 24-MAR-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9206372.6
? FILING DATE: 23-SEP-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9525004.9
? FILING DATE: 07-DEC-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9610824.6
? FILING DATE: 23-MAY-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/GB92/02240
? FILING DATE: 02-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/244,597
? FILING DATE: 01-JUN-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: David W. Clough
? REGISTRATION NUMBER: 36,107
? REFERENCE/DOCKET NUMBER: 28111/33308
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 312-474-6300
? INFORMATION FOR SEQ ID NO: 20:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 357 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
?
US-08-652-816A-20
?
Query Match 100.0%; Score 15; DB 2; Length 357;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTACTACTGAGC 15
Db 91 AGTACTACTGAGC 105
?
RESULT 3
US-08-545-809A-59
? Sequence 59, Application US/08545809A
? Patent No. 6096878
? GENERAL INFORMATION:
? APPLICANT: Honjo, Tasuku
? APPLICANT: Matsuda, Fumihiko
? TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
? TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
? NUMBER OF SEQUENCES: 145
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Fish & Richardson, P.C.
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: MA
? COUNTRY: US
? ZIP: 02110-2804

```

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/545,809A
? FILING DATE: 27-MAR-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP93/00603
? FILING DATE: 10-MAY-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Freeman, John W.
? REGISTRATION NUMBER: 29,066
? REFERENCE/DOCKET NUMBER: 06501/004001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617-542-5070
? TELEFAX: 617-542-8906
? TELEX: 200154
? INFORMATION FOR SEQ ID NO: 59:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 622 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: Genomic DNA
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
? CELL TYPE: human lymphoblast
? CELL LINE: CGM1
?
US-08-545-809A-59

```

```

Query Match 100.0%; Score 15; DB 3; Length 622;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTACTACTGAGC 15
Db 381 AGTACTACTGAGC 395

```

```

?
RESULT 4
US-08-545-809A-4
? Sequence 4, Application US/08545809A
? Patent No. 6096878
? GENERAL INFORMATION:
? APPLICANT: Honjo, Tasuku
? APPLICANT: Matsuda, Fumihiko
? TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
? TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
? NUMBER OF SEQUENCES: 145
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Fish & Richardson, P.C.
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: MA
? COUNTRY: US
? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/545,809A
? FILING DATE: 27-MAR-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP93/00603
? FILING DATE: 10-MAY-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Freeman, John W.
? REGISTRATION NUMBER: 29,066
? REFERENCE/DOCKET NUMBER: 06501/004001

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-4

Query Match 100.0%; Score 15; DB 3; Length 650;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
|||||
DB 319 AGTTACTACTGGAGC 333

RESULT 5
US-08-545-809A-61
Sequence 61, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 1212 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1

US-08-545-809A-61

Query Match 100.0%; Score 15; DB 3; Length 1212;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
|||||
DB 386 AGTTACTACTGGAGC 400

RESULT 6
US-08-477-553A-3
Sequence 3, Application US/08477553A
Patent No. 5919910
GENERAL INFORMATION:
APPLICANT: HUGHES-JONES, Nevin C
TITLE OF INVENTION: MONOCLONAL ANTIBODIES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,553A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/856,034
FILING DATE: 23-JUNE-1992
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8925590.5
FILING DATE: 13-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 007330-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-553A-3

Query Match 93.3%; Score 14; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGGAGC 15
|||||
DB 2 GTTACTACTGGAGC 15

RESULT 7
US-08-477-553A-6
Sequence 6, Application US/08477553A
Patent No. 5919910
GENERAL INFORMATION:
APPLICANT: HUGHES-JONES, Nevin C

TITLE OF INVENTION: MONOCLONAL ANTIBODIES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/477,553A
APPLICATION NUMBER: US/08/477,553A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/856,034
FILING DATE: 23-JUNE-1992
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8925590.5
FILING DATE: 13-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Neuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 007330-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
FAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-553A-6

Query Match 93.3%; Score 14; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTTACTACTGAGC 15
Db 2 GTTACTACTGAGC 15

RESULT 8
US-09-042-353-397
Sequence 397, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lomberg, Nils
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
NUMBER OF SEQUENCES: 421
TITLE OF INVENTION: Producing Heterologous Antibodies
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
FAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 397:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-397

Query Match 93.3%; Score 14; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTTACTACTGAGC 15
Db 2 GTTACTACTGAGC 15

Db 5 GTTACTACTGGAGC 18

RESULT 9
US-09-042-353-400/c
Sequence 400, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lomborg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
TITLE OF INVENTION: Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042.353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417

FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-0090400S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 400:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-400
Query Match 93.3% Score 14: DB 4: Length 43:
Best Local Similarity 100.0% Pred. No. 19:
Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
Qy 2 GTTACTACTGGAGC 15
Db 16 GTTACTACTGGAGC 3
RESULT 10
US-08-758-417A-247
Sequence 247, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lomborg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131

FILED DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 247:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 247:
US-08-758-417A-247

Query Match 93.3%; Score 14; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGGAGC 15
|||||
Db 5 GTTACTACTGGAGC 18

RESULT 11
US-08-758-417A-250/c
Sequence 250, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131

FILED DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 250:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 250:
US-08-758-417A-250

Query Match 93.3%; Score 14; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGGAGC 15
|||||
Db 16 GTTACTACTGGAGC 3

RESULT 12
US-09-042-353-148
Sequence 148, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-042-353-148

Query Match 93.3%; Score 14; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GTTACTCTGGAGC 15
DB 44 GTTACTCTGGAGC 57

RESULT 13
US-08-758-417A-412
Sequence 412, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-DEC-1996
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 412:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 412:
US-08-758-417A-412

Query Match 93.3%; Score 14; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GTTACTCTGGAGC 15
DB 44 GTTACTCTGGAGC 57

RESULT 14
US-09-042-353-149
Sequence 149, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 149:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-149

Query Match 93.3%, Score 14; DB 4; Length 282;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GTTACTACTGAGC 15
|||||
Db 44 GTTACTACTGAGC 57
RESULT 15
US-08-758-417A-413
; Sequence 413, Application US/08758417A
; Patent No. 6300129
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
; Producing Heterologous Antibodies
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,417A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Serafini, Andrew T.
; REGISTRATION NUMBER: 41,303
; REFERENCE/DOCKET NUMBER: 014643-009030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 413:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 413:
US-08-758-417A-413

Query Match 93.3%, Score 14; DB 4; Length 282;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGGAGC 15
|||||
Db 44 GTTACTACTGGAGC 57

Search completed: June 26, 2003, 04:52:44
Job time : 25.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agtactactgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	15	100.0	171	10	US-09-864-761-22522
2	15	100.0	281	10	US-09-864-761-22248
3	15	100.0	319	10	US-09-864-761-22248
4	15	100.0	378	9	US-10-067-800-67
5	15	100.0	413	9	US-09-918-995-16699
6	15	100.0	468	9	US-09-918-995-16650
7	15	100.0	462	10	US-09-864-761-5474
8	15	100.0	470	10	US-09-864-761-5760
9	15	100.0	475	10	US-09-864-761-11831
C 10	15	100.0	21470	9	US-10-092-154-1157
11	15	100.0	324	10	US-09-764-847-1157
12	15	93.3	341	9	US-09-864-761-31244
13	15	93.3	341	9	US-09-879-813-15
14	15	93.3	341	9	US-10-146-505-15
15	15	93.3	353	10	US-09-864-761-28159
C 16	15	93.3	398	10	US-09-864-761-10457
17	15	93.3	417	10	US-09-905-243-7
18	15	93.3	470	10	US-09-864-761-3704
19	15	93.3	471	9	US-09-187-693-31

20	14	93.3	481	10	US-09-864-761-5338	Sequence 5338, Ap
21	14	93.3	481	10	US-09-864-761-11572	Sequence 11572, A
22	14	93.3	482	10	US-09-864-761-14715	Sequence 14715, A
23	14	93.3	505	10	US-09-954-456-1183	Sequence 1183, Ap
24	14	93.3	669	9	US-09-972-656-65	Sequence 65, Appl
25	14	93.3	687	9	US-09-972-656-81	Sequence 81, Appl
26	14	93.3	690	9	US-09-972-656-71	Sequence 71, Appl
C 27	14	93.3	2624	9	US-09-854-133-19	Sequence 19, Appl
28	14	93.3	2624	9	US-09-738-973-19	Sequence 19, Appl
C 29	13.4	89.3	303	10	US-09-783-590-7401	Sequence 7401, Ap
30	13.4	89.3	320	9	US-09-764-891-9777	Sequence 9777, Ap
31	13.4	89.3	342	10	US-09-828-708-121	Sequence 121, App
32	13.4	89.3	362	10	US-09-864-761-32297	Sequence 32297, A
33	13.4	89.3	362	10	US-09-878-574-1765	Sequence 1765, Ap
34	13.4	89.3	363	9	US-10-067-800-59	Sequence 59, Appl
35	13.4	89.3	456	9	US-09-187-693-32	Sequence 32, Appl
C 36	13.4	89.3	466	9	US-10-198-846-7483	Sequence 7483, Ap
37	13.4	89.3	469	10	US-09-864-761-15792	Sequence 15792, A
38	13.4	89.3	472	10	US-10-040-739-963	Sequence 963, App
C 39	13.4	89.3	474	10	US-09-867-701-9902	Sequence 9902, Ap
40	13.4	89.3	510	9	US-10-198-846-3994	Sequence 3994, Ap
41	13.4	89.3	593	9	US-09-764-868-192	Sequence 192, App
C 42	13.4	89.3	831	10	US-09-827-663-3	Sequence 3, Appl1
C 43	13.4	89.3	1278	10	US-09-815-242-9165	Sequence 9165, Ap
C 44	13.4	89.3	1290	10	US-09-815-242-9491	Sequence 9491, Ap
C 45	13.4	89.3	1293	10	US-09-827-663-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-864-761-22522/c
Sequence 22522, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemolca-X-1
CURRENT APPLICATION NUMBER: US/09/864, 761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180, 312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207, 456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632, 366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263, 6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236, 359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

```

? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 22522
? LENGTH: 171
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC004593.1
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
? OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 28
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
? OTHER INFORMATION: NT HIT: 918567391, EVALUATE 1.60e+00
? OTHER INFORMATION: EST_HUMAN HIT: BE27658.1, EVALUATE 1.50e-01
? OTHER INFORMATION: SWISSPROT HIT: P07131, EVALUATE 7.80e-01
US-09-864-761-22522

Query Match      100.0%; Score 15; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
      |||||||
Db      82 AGTTACTACTGAGC 68

RESULT 2
US-09-864-761-22248
? Sequence 22248, Application US/09864761
? Patent No. US20020048763A1
? GENERAL INFORMATION:
? APPLICANT: Penn, Sharon G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? APPLICANT: Chen, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
? FILE REFERENCE: Aeomica-X-1
? CURRENT APPLICATION NUMBER: US/09/864,761
? CURRENT FILING DATE: 2001-05-23
? PRIOR APPLICATION NUMBER: US 60/180,312
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: US 60/207,456
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```

? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 22248
? LENGTH: 281
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AB019441.1
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
? OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 5.7
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.9
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.3
? OTHER INFORMATION: EST_HUMAN HIT: AM405472.1, EVALUATE 0.00e+00
? OTHER INFORMATION: NT HIT: Z14240.1, EVALUATE 0.00e+00
? OTHER INFORMATION: SWISSPROT HIT: P06331, EVALUATE 2.00e-30
US-09-864-761-22248

Query Match      100.0%; Score 15; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
      |||||||
Db      64 AGTTACTACTGAGC 78

RESULT 3
US-09-864-761-28401
? Sequence 28401, Application US/09864761
? Patent No. US20020048763A1
? GENERAL INFORMATION:
? APPLICANT: Penn, Sharon G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? APPLICANT: Chen, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
? FILE REFERENCE: Aeomica-X-1
? CURRENT APPLICATION NUMBER: US/09/864,761
? CURRENT FILING DATE: 2001-05-23
? PRIOR APPLICATION NUMBER: US 60/180,312
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: US 60/207,456
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
```



```

; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28401
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB019438.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56
; OTHER INFORMATION: NT HIT: U03896.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW402332.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P06331, EVALUE 3.00e-36
; US-09-864-761-28401

Query Match      100.0%; Score 15; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||
Db      102 AGTTACTACTGAGC 116

RESULT 4
US-10-067-800-67
; Sequence 67, Application US/10067800
; Publication No. US20030100058A1
; GENERAL INFORMATION:
; APPLICANT: Roschke, Viktor
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven, M.
; TITLE OF INVENTION: Human G-protein Chemokine Receptor (CCR5) HDGNR10
; FILE REFERENCE: 1488.1150001
; CURRENT APPLICATION NUMBER: US/10/067,800
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: PCT/US01/04153
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/779,880
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/297,257
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/310,458
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/328,447
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/341,725
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 378
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(378)
; OTHER INFORMATION:
; US-10-067-800-67

Query Match      100.0%; Score 15; DB 9; Length 378;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||
Db      91 AGTTACTACTGAGC 105

RESULT 5
US-09-918-995-16699
; Sequence 16699, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16699
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(413)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-16699

Query Match      100.0%; Score 15; DB 9; Length 413;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||
Db      171 AGTTACTACTGAGC 185

RESULT 6
US-09-918-995-16650
; Sequence 16650, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16650
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-918-995-16650

Query Match      100.0%; Score 15; DB 9; Length 438;
```

Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0;

Indels 0;

Gaps 0;

QY 1 AGTACTACTGAGC 15
Db 166 AGTACTACTGAGC 180

RESULT 7

US-09-864-761-5474
Sequence 5474, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 5474
LENGTH: 462
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AB019441.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.9

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.3
US-09-864-761-5474

Query Match 100.0%; Score 15; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTACTACTGAGC 15
Db 316 AGTACTACTGAGC 330

RESULT 8

US-09-864-761-5760/C
Sequence 5760, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 5760
LENGTH: 470
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004593.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 28

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
US-09-864-761-5760

Query Match 100.0%; Score 15; DB 10; Length 470;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
|||||
DB 358 AGTTACTACTGGAGC 344

RESULT 9
US-09-864-761-11831
Sequence 11831, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/532,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11831
LENGTH: 475
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO AB019438.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56
US-09-864-761-11831

Query Match 100.0%; Score 15; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
|||||
DB 329 AGTTACTACTGGAGC 343

RESULT 10
US-10-092-154-1157/c
Sequence 1157, Application US/10092154
Publication No. US20030054375A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009C1
CURRENT APPLICATION NUMBER: US/10/092,154
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2003
Prior Application removed - See File Wrapper or Palm
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1157
LENGTH: 21470
TYPE: DNA
ORGANISM: Homo sapiens
US-10-092-154-1157

Query Match 100.0%; Score 15; DB 9; Length 21470;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
|||||
DB 10965 AGTTACTACTGGAGC 10951

RESULT 11
US-09-764-847-1157/c
Sequence 1157, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 2003
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1157
LENGTH: 21470
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-847-1157

Query Match 100.0%; Score 15; DB 10; Length 21470;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
|||||
DB 10965 AGTTACTACTGGAGC 10951

RESULT 12
US-09-864-761-31244
Sequence 31244, Application US/09864761
Patent No. US20020048763A1

```

: GENERAL INFORMATION:
: APPLICANT: Sale, Julian E.
: APPLICANT: Neuberger, Michael S.
: APPLICANT: Cumbers, Sarah J.
: TITLE OF INVENTION: Method of Generating Diversity
: FILE REFERENCE: 18396/2005
: CURRENT APPLICATION NUMBER: US/09/879,813
: CURRENT FILING DATE: 2001-06-11
: PRIOR APPLICATION NUMBER: 09/828,717
: PRIOR FILING DATE: 2001-06-04
: PRIOR APPLICATION NUMBER: PCT/GB99/03358
: PRIOR FILING DATE: 1999-10-08
: NUMBER OF SEQ ID NOS: 87
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 15
: LENGTH: 341
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(341)
: OTHER INFORMATION:
US-09-879-813-15

Query Match          93.3%: Score 14; DB 9; Length 341;
Best Local Similarity 100.0%: Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy          2 GTTACTACTGAGC 15
            |||
Db          74 GTTACTACTGAGC 87

RESULT 14
US-10-146-505-15
: Sequence 15, Application US/10146505
: Publication No. US20030108889A1
: GENERAL INFORMATION:
: APPLICANT: Sale, Julian E.
: APPLICANT: Neuberger, Michael S.
: APPLICANT: Cumbers, Sarah J.
: TITLE OF INVENTION: Method of Generating Diversity
: FILE REFERENCE: 18396/2005B
: CURRENT APPLICATION NUMBER: US/10/146,505
: CURRENT FILING DATE: 2002-11-18
: PRIOR APPLICATION NUMBER: 09/828,717
: PRIOR FILING DATE: 2001-04-06
: PRIOR APPLICATION NUMBER: 09/879,813
: PRIOR FILING DATE: 2001-06-11
: PRIOR APPLICATION NUMBER: PCT/GB99/03358
: PRIOR FILING DATE: 1999-10-08
: PRIOR APPLICATION NUMBER: GB 9822104.7
: PRIOR FILING DATE: 1998-10-09
: PRIOR APPLICATION NUMBER: GB 9901141.3
: PRIOR FILING DATE: 1999-01-19
: PRIOR APPLICATION NUMBER: GB 9913435.5
: PRIOR FILING DATE: 1999-06-09
: NUMBER OF SEQ ID NOS: 127
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 15
: LENGTH: 341
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(341)
: OTHER INFORMATION:
US-10-146-505-15

```

OY 2 GTTACTACTGAGC 15
 |||||
 Db 74 GTTACTACTGAGC 87

RESULT 15

US-09-864-761-28159
 ; Sequence 28159, Application US/09864761
 ; Patent No. US2002048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aeomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 28159
 ; LENGTH: 353
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AB019439.1
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.54
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.57
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56
 ; OTHER INFORMATION: NT HIT: X92278.1, EVALU0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: P06331, EVALU0.00e+00
 ; OTHER INFORMATION: EST_HUMAN HIT: BE672445.1, EVALU0.00e+00
 ; US-09-864-761-28159

Query Match 93.3%; Score 14; DB 10; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 GTTACTACTGAGC 15
 |||||
 Db 103 GTTACTACTGAGC 116

Search completed: June 26, 2003, 04:57:35
 Job time : 61.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agtactactgagc 15

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	187	10	AM401419 UI-HF-BK0
2	15	100.0	248	10	AM403144 UI-HF-BK0
3	15	100.0	273	12	BF847860 IL5-EN008
4	15	100.0	275	10	AM408669 UI-HF-BM0
5	15	100.0	299	14	T27727 EST13874 HU
6	15	100.0	319	10	AM405472 UI-HF-BL0

7	15	100.0	367	10	AM403544 UI-HF-BK0
8	15	100.0	421	10	AM407630 UI-HF-BM0
9	15	100.0	459	10	AM402533 UI-HF-BK0
10	15	100.0	465	10	AM402332 UI-HF-BK0
11	15	100.0	496	10	AM402282 UI-HF-BK0
12	15	100.0	539	10	AM786572 119835 MA
13	15	100.0	548	10	AM402602 UI-HF-BK0
14	15	100.0	590	10	AV983832 AV983832
15	15	100.0	649	12	BM783015 K-EST0060
16	15	100.0	667	14	BM757507 602714665
17	15	100.0	661	12	BG686421 602638356
18	15	100.0	680	12	BG684306 602636046
19	15	100.0	729	13	BI488497 603021006
20	15	100.0	733	13	BM008087 603617861
21	15	100.0	736	12	BG685592 602637569
22	15	100.0	790	12	BG754804 602714380
23	15	100.0	791	12	BG685529 602637396
24	15	100.0	820	13	BI770035 603053578
25	15	100.0	844	13	BI489640 603032108
26	15	100.0	850	12	BG686474 602637417
27	15	100.0	911	14	BQ711708 60710745
28	15	100.0	918	14	BQ708269 60710745
29	15	100.0	920	14	BQ710745 60710745
30	15	100.0	932	14	BQ710283 60710283
31	15	100.0	939	9	AL541900 602438679
32	15	100.0	947	14	BQ709527 60709527
33	15	100.0	951	14	BQ709739 60709739
34	15	100.0	964	12	BG757625 602711208
35	15	100.0	971	9	AL560682 602438679
36	15	100.0	977	12	BG397634 602438679
37	15	100.0	980	12	BG684861 602636587
38	15	100.0	1005	14	BQ709036 60709036
39	15	100.0	1036	14	BQ072430 6072430
40	15	100.0	1108	14	BM920469 602469
41	15	100.0	1118	14	BO882037 60882037
42	14	93.3	113	17	AZ013956 6013956
43	14	93.3	186	14	BQ354864 60354864
44	14	93.3	196	10	AM401400 6024100
45	14	93.3	214	13	BI225563 602948847

ALIGNMENTS

RESULT 1
LOCUS AM401419 187 bp mRNA linear EST 16-FEB-2000
DEFINITION UI-HF-BK0-aaa-e-11-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3053276 5', mRNA sequence.

ACCESSION AM401419
VERSION AM401419.1 GI:6920201
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 187)
AUTHORS Eukariyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE NIH-MGC http://mgc.ncl.nih.gov/
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
CDNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www.bio.lnl.gov/bhrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers 1..187

FEATURES
Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3053276"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_line="germlinal center B cells"
/lab_host="DH10B (LTI)"
/Note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      34 a      54 c      61 g      38 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 8,4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGGAGC 15
        |||||||
Db      122 AGTTACTACTGGAGC 136

RESULT 2
AM403144      248 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UT-HF-BKO-aar-g-10-0-UT.r1 NIH_MGC_36 Homo sapiens cDNA clone
DEFINITION      IMAGE:3054907 5', mRNA sequence.
ACCESSION      AM403144
VERSION      AM403144.1 GI:6921997
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 248)
NIH-MGC http://mhc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bdrp/image/image.html
Seq primer: M13 Forward.
FEATURES
source
location/Qualifiers
1..248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3054907"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_line="germlinal center B cells"
/lab_host="DH10B (LTI)"
/Note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      55 a      73 c      69 g      51 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 9,4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AGTTACTACTGGAGC 15
        |||||||
Db      115 AGTTACTACTGGAGC 129

RESULT 3
BF847860/c      273 bp      mRNA      linear      EST 16-JAN-2001
LOCUS      IL5-EN0085-021100-228-d07 EN0085 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BF847860
ACCESSION      BF847860
VERSION      BF847860.1 GI:12235023
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 273)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R.,
Negal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?IL5&IL5-EN0085-
021100-228-d07&IL5-2000-11-02&IL5-1)
Seq primer: puc 18 forward
High quality sequence stop: 272.
FEATURES
source
location/Qualifiers
1..273
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0085"
/dev_stage="Adult"
/Note="Organ: lung, normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from OHSR5 PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      61 a      76 c      81 g      55 t
ORIGIN

Query Match      100.0%; Score 15; DB 12; Length 273;
Best Local Similarity 100.0%; Pred. No. 9,7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGGAGC 15
        |||||||
Db      175 AGTTACTACTGGAGC 161

RESULT 4
AM408669      275 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UT-HF-BKO-ads-b-08-0-UT.r1 NIH_MGC_38 Homo sapiens cDNA clone
DEFINITION      IMAGE:3062726 5', mRNA sequence.
ACCESSION      AM408669

```


VERSION AM408669.1 GI:6927726
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 275)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: M.B. Soares Lab
 CDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 forward.
 FEATURES
 source Location/Qualifiers
 1. 275
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3062726"
 /clone_lib="NIH_MGC_38"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTR)"
 /note="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (2.5-3.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
 BASE COUNT 62 a 84 c 73 g 56 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 275;
 Best Local Similarity 100.0%; Pred. No. 9.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AGTACTACTGAGC 15
 |||||||||||||
 Db 105 AGTACTACTGAGC 119
 RESULT 5
 T27727 299 bp mRNA linear EST 06-SEP-1995
 LOCUS EST13874 Human Testis Homo sapiens CDNA 5' end similar to
 DEFINITION Immunoglobulin mu (gamma) heavy chain, V(IV)DTC regions (HT:3057),
 mRNA sequence.
 T27727
 T27727.1 GI:609825
 EST.
 human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 299)
 AUTHORS Adams M.D., Kerlavage A.R., Fleschmann R.D., Fulner R.A., Bult C.J., Lee N., Kirness E.F., Weinstein K.G., Gocayne J.D., White O., Sutton G., Blake J.A., Brandon R.C., Chiu M.-W., Clayton R.A., Cline R.T., Cotton M.D., Earle-Hughes J., Fine L.D., Fitzgerald L.M., FitzHugh W.M., Fritchman V.L., Geoghegan N.S.M., Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkley P.S., Kelley J.M., Klimchik K.M., Kelley J.C., Liu L.-I., Marmatos S.M., Merrick J.M., Moreno-Palances R.F., McDonald L.A., Nguyen D.T., Pellegrino S.M., Phillips C.A., Ryder S.E., Scott J.L., Saudak D.M., Shirley R., Small K.V., Spriggs T.A., Uterback T.R., Weidman J.F., Li Y., Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.-J.,

TITLE
 JOURNAL Nature 377, 3-174 (1995)
 MEDLINE 96026280
 COMMENT Contact: Venter, JC
 The Institute for Genomic Research
 932 Clopper Rd, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: tblinfo@tdb.tigr.org
 For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tblinfo@tdb.tigr.org)
 Seq primer: M13 Reverse.
 FEATURES
 source Location/Qualifiers
 1. 299
 /organism="Homo sapiens"
 /db_xref="AFCC (lnhost):108071"
 /db_xref="taxon:9606"
 /clone_lib="Human Testis"
 /note="Organ: testis"
 BASE COUNT 61 a 84 c 82 g 68 t 4 others
 ORIGIN
 Query Match 100.0%; Score 15; DB 14; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AGTACTACTGAGC 15
 |||||||||||||
 Db 169 AGTACTACTGAGC 183
 RESULT 6
 AM405472 319 bp mRNA linear EST 16-FEB-2000
 LOCUS UT-HF-BL0-adh-d-02-0-UT.NIH_MGC_37 Homo sapiens CDNA clone
 DEFINITION IMAGE:3061658 5', mRNA sequence.
 AM405472
 AM405472.1 GI:6924529
 EST.
 human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 319)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: M.B. Soares Lab
 CDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 forward.
 FEATURES
 source Location/Qualifiers
 1. 319
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3061658"
 /clone_lib="NIH_MGC_37"
 /tissue_type="lymph"

```

/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(1.5-2.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldio, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      74 a      92 c      93 g      60 t
ORIGIN
Query Match      100.0%; Score 15; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
|||||
Db 100 AGTTACTACTGGAGC 114

```

```

RESULT 7
LOCUS      AM403544      367 bp      mRNA      linear      EST 16-FEB-2000
DEFINITION UI-HF-BKO-abf-d-09-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3055937 5', mRNA sequence.
ACCESSION  AM403544
VERSION     AM403544.1 GI:6922517
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 367)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: M.B. Soares Lab
            CDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bdrip/image/image.html
            Seq primer: M13 Forward.
FEATURES
    source
        Location/Qualifiers
            1..367
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3055937"
                /clone_lib="NIH_MGC_36"
                /tissue_type="lymph"
                /cell_type="germinal center B cells"
                /cell_line="MGC85"
                /lab_host="DH10B (LTI)"
                /note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
                Constructed from size fractionated cytoplasmic mRNA
                (0.5-1.5kb). Directionally cloned. Cells provided by Louis
                M. Staudt, Ph.D. Library preparation by Maria de Fatima
                Bonaldio, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      80 a      104 c      109 g      73 t      1 others
ORIGIN
Query Match      100.0%; Score 15; DB 10; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
|||||
Db 115 AGTTACTACTGGAGC 129

```

```

RESULT 8
LOCUS      AM407630      421 bp      mRNA      linear      EST 16-FEB-2000
DEFINITION UI-HF-BKO-abd-g-a-10-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
IMAGE:3062322 5', mRNA sequence.
ACCESSION  AM407630
VERSION     AM407630.1 GI:6926697
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 421)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: M.B. Soares Lab
            CDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bdrip/image/image.html
            Seq primer: M13 Forward.
FEATURES
    source
        Location/Qualifiers
            1..421
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3062322"
                /clone_lib="NIH_MGC_38"
                /tissue_type="lymph"
                /cell_type="germinal center B cells"
                /cell_line="MGC85"
                /lab_host="DH10B (LTI)"
                /note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
                Constructed from size fractionated cytoplasmic mRNA
                (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                M. Staudt, Ph.D. Library preparation by Maria de Fatima
                Bonaldio, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      91 a      131 c      111 g      87 t      1 others
ORIGIN
Query Match      100.0%; Score 15; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
|||||
Db 93 AGTTACTACTGGAGC 107

```

```

RESULT 9
LOCUS      AM402533      459 bp      mRNA      linear      EST 16-FEB-2000
DEFINITION UI-HF-BKO-aas-f-08-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3054878 5', mRNA sequence.
ACCESSION  AM402533
VERSION     AM402533.1 GI:6921226
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 459)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov

```

Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab
 cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bdrp/image/image.html
 Seq primer: M13 Forward.

FEATURES

source

1..459

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3054878"

/clone_1lb="NIH_MGC_36"

/tissue_type="lymph"

/cell_type="germinal center B cells"

/lab_host="DH10B (LTI)"

/note="vector: p7773-Pac; Site_1: NotI; Site_2: Eco RI;

Constructed from size fractionated cytoplasmic mRNA

(0.5-1.5kb). Directionally cloned. Cells provided by Louis

M. Staudt, Ph.D. Library preparation by Maria de Fatima

Bonaldio, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT

95 a 148 c 124 g 92 t

ORIGIN

Query Match

100.0%; Score 15; DB 10; Length 459;

Best Local Similarity 100.0%; Pred. NO. 1.2e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15

Db 87 AGTTACTACTGGAGC 101

Db

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

87 AGTTACTACTGGAGC 101

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3053915"

/clone_1lb="NIH_MGC_36"

/tissue_type="lymph"

/cell_type="germinal center B cells"

/lab_host="DH10B (LTI)"

/lab_host="DH10B (LTI)"

/lab_host="DH10B (LTI)"

/lab_host="DH10B (LTI)"

BASE COUNT

102 a 137 c 146 g 80 t

ORIGIN

Query Match

100.0%; Score 15; DB 10; Length 465;

Best Local Similarity 100.0%; Pred. NO. 1.2e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15

Db 122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

122 AGTTACTACTGGAGC 136

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3054035"

/clone_1lb="NIH_MGC_36"

/tissue_type="lymph"

/cell_type="germinal center B cells"

/lab_host="DH10B (LTI)"

/note="vector: p7773-Pac; Site_1: NotI; Site_2: Eco RI;

Constructed from size fractionated cytoplasmic mRNA

(0.5-1.5kb). Directionally cloned. Cells provided by Louis

M. Staudt, Ph.D. Library preparation by Maria de Fatima

Bonaldio, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

Bonalido, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT

102 a 160 c 123 g 111 t

ORIGIN

Query Match

100.0%; Score 15; DB 10; Length 496;

Best Local Similarity 100.0%; Pred. NO. 1.2e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15

Db 53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

53 AGTTACTACTGGAGC 67

LOCUS AW786572 539 bp mRNA linear EST 09-JUL-2000
 DEFINITION 119835 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION AW786572
 VERSION AW786572.1 GI:7843348
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 1 (bases 1 to 539)
 Fahnenkrug, S.C., Fieking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.,
 and Keefe, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt. trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCCAGTCAGCAGC
 Plate: 42 row: D column: 18
 Seq primer: ATTTAGTGACATATAG.
 Location/Qualifiers
 1..539
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 1Pig"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."
 BASE COUNT 164 a 101 c 142 g 132 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 539;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTTACTACTGGAGC 15
 ||||||||||||||||
 Db 348 AGTTACTACTGGAGC 334
 ||||||||||||||||
 RESULT 13 548 bp mRNA linear EST 16-FEB-2000
 LOCUS AW402602
 DEFINITION UT-HF-BK0-aay-f-06-0-ut-rl NIH_MGC_36 Homo sapiens cDNA clone
 IMAGE:3055259 5', mRNA sequence.
 ACCESSION AW402602
 VERSION AW402602.1 GI:6921304
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 548)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-ri@mail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab

CDNA library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN. at:
www.bio.lim.gov/bdrp/image/image.html
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..548
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3055259"
 /clone_lib="NIH_MGC_36"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /lab_host="DH10B (LT1)"
 /note="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (0.5-1.5kb). Directionally cloned. Cells provided by Louis
 M. Staudt, Ph.D. Library preparation by Maria de Fatima
 Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
 BASE COUNT 115 a 171 c 148 g 114 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 548;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTTACTACTGGAGC 15
 ||||||||||||||||
 Db 109 AGTTACTACTGGAGC 123
 ||||||||||||||||
 RESULT 14 590 bp mRNA linear EST 14-MAR-2002
 LOCUS AV983832/c
 DEFINITION AV983832 Nori Satoh unpublished cDNA library, egg Clona
 intestinalis cDNA clone c1eg28e23 5', mRNA sequence.
 ACCESSION AV983832
 VERSION AV983832.1 GI:19472913
 KEYWORDS EST.
 SOURCE Clona intestinalis.
 ORGANISM Clona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Cloniidae; Clona.
 1 (bases 1 to 590)
 Satoh, N., Satou, Y., Kohara, Y. and Shin-i, T.
 Expressed genes in Clona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 Location/Qualifiers
 1..590
 /organism="Clona intestinalis"
 /db_xref="taxon:7719"
 /clone="c1eg28e23"
 /clone_lib="Nori Satoh unpublished cDNA library, egg"
 /tissue_type="whole animal"
 /dev-stage="egg"
 /note="Vector: pBluescript SK"
 BASE COUNT 200 a 152 c 101 g 137 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 590;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTTACTACTGGAGC 15

Db 243 AGTTACTACTGGAGC 229

RESULT 15

BM783015 607 bp mRNA linear EST 05-MAR-2002
K-EST0060735 S18N669761 Homo sapiens cDNA clone S18N669761-2-ED6

LOCUS 5', mRNA sequence.

ACCESSION BM783015 GI:19131247

VERSION EST.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 607)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribd.re.kr
Plate: 2 row: E column: 06
High quality sequence stop: 607.

FEATURES

source

Location/Qualifiers

1..607
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S18N669761-2-ED6"
/clone_1lb="S18N669761"
/sex="F"
/lab_host="Top10F"

/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 131 a 187 c 163 g 126 t

ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 607;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGGAGC 15

Db 193 AGTTACTACTGGAGC 207

Search completed: June 26, 2003, 04:50:22
Job time : 753.2 secs

